SHORT TO MEDIUM RANGE HYDROMETEOROLOGICAL FORECASTING IN THE RIO GRIJALVA BASIN, MEXICO

by

Edgar Misael Uribe Alcantara

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF HYDROLOGY AND WATER RESOURCES

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY
WITH A MAJOR IN HYDROLOGY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2007

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation

prepared by Edgar Misael Uribe Alcantara

entitled Short to Medium Range Hydrometeorological Forecasting in the Rio Grijalva Basin, Mexico

and recommend that it be accepted as fulfilling the dissertation requirement for the

Degree of Doctor of Philosophy

_____ Date: 5/9/07
William James Shuttleworth

_____ Date: 5/9/07
Hoshin V. Gupta

_____ Date: 5/9/07
Steven L. Mullen

_____ Date: 5/9/07
Xubin Zeng

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

_____ Date: 5/9/07
Dissertation Director: William James Shuttleworth

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

SIGNED: Edgar Misael Uribe Alcantara

ACKNOWLEDGEMENTS

I would like to acknowledge the financial support from CONACyT through its program of scholarships for graduate students. I also received support under NOAA Grant NA16GP2002 and from the SAHRA Center under Cooperative Agreement EAR-9876800 and NSF award DEB–0415977.

My gratitude goes to the members of my dissertation committee: William J. Shuttleworth, Hoshin V. Gupta, Steven L. Mullen, and Xubin Zeng. My deepest acknowledgement goes to “Jim” for his patience, support and generosity, despite the countless difficulties we faced together along the way. I’ve always believed people who excel in their field can also be sympathetic, humble, and generous. You certainly are a living example of this conviction. A strong feeling of gratitude goes to Hoshin. Shy people tend to erase the phrase “I think..” from their language but you always helped me found and shape my own ideas so that I would feel inspired and confident enough to pursue them. I would like to thank Steve and Xubin for their invaluable feedback, and for showing a strong disposition to meet for examinations and discussions.

I would like to thank faculty, staff, and colleagues from HWR and the UofA: Jim Washburne for his consistent stewardship and practical support; Bart Nijssen for guiding me in the selection of the hydrological model; Matej Durcik for sharing his code for PERSIANN; Koray Yilmaz for his tips with Matlab; Martha Conklin for trusting me as her TA; James Broermann, Dean Jones, and Nate Bryant, for their computational wizardry; Terrie Thompson, Marissa Rodriguez, Melissa Rodriguez, Rannie Fox, Annica Polm, Tomas Alvarez, Wanda Shirey, Erma Santander, Maria Teresa Velez, Paul Pronze, and Blanche Swyers, for finding successful paths in my labyrinth-like life and for being an example of professionalism and happy commitment; Jesus Gastelum, Eric Morales, Demetrio Fernandez, Erick Sanchez, Lenom Cajuste, and Rafael Rosolem, for their patient ears and sensible advice.

Thanks also to the following people from the USGS for providing and guiding me in the use of MMS-PRMS, GIS-Weasel, and LUCA: George Leavesley, Steve Markstrom, Roland Viger, Lauren Hay, and Makiko Umemoto and to our colleagues in the University of Irvine for their support with the PERSIANN dataset: Xiaogang Gao, Dan Braithwaite, and K. Hsu.

I would like to thank my colleagues from the CCA-UNAM: Víctor Magaña and Ernesto Caetano for delivering me in good hands after I finished my master’s degree, and for being the catalyst of this research; Jorge Luis Vazquez for his support with the meteorological networks; and Cecilia Conde, Carlos Gay, Baldemar Mendez, Gustavo Vazquez, Matias Mendez and Heidi Ayala. I also thank my colleagues from IMTA José Luis Perez, Ricardo Prieto, José A. Salinas, María de los Angeles Suárez Medina, and René Lobato for motivating this research. Thanks to David Gochis and Chris Watts for letting me participate in the NAM project. My most sincere apologies if I have forgotten anyone else.

DEDICATION

A mis amores: Yeni, Natalia, Natividad, José, Alfredo, Marisol, y humildemente a Dios con profundo agradecimiento por las segundas oportunidades.

TABLE OF CONTENTS

TABLE OF CONTENTS - *Continued*

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS - *Continued*

LIST OF ILLUSTRATIONS - *Continued*

LIST OF TABLES

ABSTRACT

The Rio Grijalva basin is the most important basin in Mexico in terms of hydropower production and damages related to extreme rainfall events. This study investigates establishing a short- to medium-range hydrometeorological forecasting system for this basin which comprises a hydrological model and a regional Numerical Weather Prediction Model (NWPM). A physical, distributed, hydrological model (MMS-PRMS) is implemented through the following steps: (1) basin parameterization; (2) interfacing to observed meteorological fields, and (3) parameter optimization. The datasets normally used to parameterize the MMS-PRMS are only available in the US so an alternative methodology for deriving parameters from globally available public datasets was devised. Modeled streamflow calculated by model with the initial parameters was in good agreement with observed streamflow, and optimization yielded even better agreement. The predictive capabilities of the hydrological model was then tested by implementing modeled rainfall and temperature from the *North American Regional Reanalysis* (NARR), these data being used as a surrogate for those that would be available from a regional NWPM. A significant bias in NARR-rainfall was identified and a novel probabilistic correction procedure devised. This procedure was then extended to provide estimates of uncertainty in the modeled streamflow. Within the calculated uncertainty, the modeled streamflow calculated with these corrected NARR data is in good agreement with modeled streamflow calculated using local meteorological data.

1. INTRODUCTION

1.1. Background and Motivation

Forecasting precipitation, along with temperature and wind, has historically been one of the primary objectives of meteorology because of its relevance to human activities. It is in fact one of the main motivations behind the development of Numerical Weather Prediction Models (NWPMs), which are perhaps the most complex and complete compilations of meteorological understanding, and which are specifically focused on prediction.

A recent study of the improvement of three global NWPMs over the last decade suggests that their skill has increased significantly (Simmons and Hollingsworth, 2002). One of the factors instrumental in this improvement is the advances that have been made in computer technology, which has allowed the generation of forecast ensembles operationally. Ensembles involve many model realizations with slightly different initial conditions to account for uncertainties in the data assimilation process used during model initiation, and the resulting range of realizations may give some indication of the forecast spread. The European Centre for Medium Range Weather Forecasting (ECMWF) Ensemble Prediction System, for example, provides skilful prediction of low precipitation amounts in forecasts of up to six days ahead, and of high precipitation amounts in forecasts of up to four days ahead (Buizza and Palmer, 1995). Severe storms are not modeled as successfully yet, but it has been shown that the ECMWF Ensemble Prediction System can provide early indications of possible severe

storm occurrences (Buizza and Hollingsworth, 2002). Ensemble forecasts of large scale atmospheric flow patterns have also positive skill scores with respect to deterministic forecasts for up to 10 days ahead (Chessa and Lalaurette, 2001); and estimates of precipitation quantity, timing, and spatial distribution can be made up to 10 days ahead for model grid scales of 40 km in deterministic mode, and 80 km in ensemble mode.

As the accuracy, resolutions, and leading times of meteorological forecasts increase, the range of applications also increases. A promising potential application would be to combine meteorological models and hydrologic models to provide forecasts of variables such as streamflow and soil moisture, whose variation can then be directly related to, for example, flood occurrence, hydropower production, and water resources. However, the combination of hydrological models and global NWPMs is not straightforward. The scales used in global NWPMs are too coarse for the estimation of fields over most river basins. Therefore, a downscaling approach in which global forecasts are used as boundary conditions for regional NWPMs that then generate forecasts with higher resolution has frequently been explored (De Roo et al., 2003;Westrick et al., 2002;Westrick and Mass, 2001). This approach is usually attempted using regional NWPMs which are operational and well established and supported by a professional team of experienced specialists.

However, in Mexico, operational forecasts based on NWPMs are still in the early stages of development. The institution officially responsible for providing

weather forecasts is the *Servicio Meteorológico Nacional* (National Meteorological Service, SMN). For a long time the weather forecasts provided by this institution were based only on synoptic analysis, but two years ago the SMN started to provide deterministic forecasts based on NWPMs to the public (http://smn.cna.gob.mx/). A second significant effort in this respect is the regional NWPM for Mexico and Mexico City of the *Centro de Ciencias de la Atmósfera de la Universidad Nacional Autónoma de México* (Center for Atmospheric Sciences from the National Autonomous University of Mexico, CCA-UNAM). This system has been running operationally for about five years (Pérez, 2004). Although the CCA-UNAM is not formally responsible for providing weather forecasts, and the interest is mainly scientific, the results are published electronically (http://pembu.atmosfcu.unam.mx/version/index.html).

Thus there is some growing skill in using NWPMs in Mexico, but currently the use of hydrological models by scientific and governmental institutions in Mexico is practically none existent. Some exceptions include analysis of the value of rainfall estimations from radar when implemented into an empirical, event-based, hydrologic model (Baldemar-Mendez, 2005), and attempts which solely rely on using statistical relationships between measured rainfall and runoff. Among the important reasons why there is the limited use of hydrological models in Mexico is the limited number of relevant specialists, and the lack of models that are appropriate for Mexican basins, the latter being largely related to the lack of proper documentation from which hydrological models can be implemented.

The national datasets which describe vegetation cover and soils published by the *Instituto Nacional de Estadística Geografía e Informática* (National Institute of Statistics, Geography and Informatics, INEGI) are mostly descriptive and therefore an unsuitable basis for physically-based hydrological modeling.

Physically-based, distributed, continuous hydrological models (sometimes referred to below as “complex” models) usually have a higher information requirement for parameterization than empirical, lumped, event-based models (sometimes referred to below as “simple” models). The number of parameters required can be large, especially if the model is distributed, thus increasing the difficulty of using parameter optimization techniques. Therefore using simple models might seem more appropriate for Mexican river basins. However, the recent advent of novel and powerful optimization techniques, along with the development of freely-available global databases based on satellite observations, should facilitate the implementation of complex hydrological models in poorly documented basins. Exploring the potential to do this was an important component of this study.

It is to be anticipated that complex hydrological models will be better able to take full-advantage of the increasing forecast lead-time and resolution that is resulting from the continuing development of NWPMs, and that they will generate model products with similar valuable characteristics. Their distributed nature means they are likely better able to acknowledge spatial variability in any estimated meteorological fields forcing derived from radar and satellites, and

represent spatial variability in predicted fields. Because they provide continuous simulation, such models may be able to calculate reasonably realistic initial and antecedent conditions through optimization and spin-up which would otherwise need to be measured. Moreover, continuous models do not necessarily require calibration for every event, unlike event-based models which therefore have less predictive use. The physical aspect of complex models also facilitates the implementation of additional hydrological processes, such as infiltration and evaporation, which can improve understanding within the basin. Therefore, if the initial issues associated with parameterization and optimization can be successfully addressed, complex models could actually be more convenient than simple models, and their products likely more accurate because the model more closely reflects current hydrological and meteorological understanding.

1.2. The Río Grijalva Basin

This Rio Grijalva basin lies mainly within the states of Chiapas and Tabasco in southeast Mexico, but partly in western Guatemala. Maps and information on the basin are limited and somewhat contradictory. In this study the information used is solely that provided officially by the *Comisión Nacional del Agua* (National Water Commission, (CONAGUA, 2006)) unless otherwise stated. CONAGUA has divided the country into several Hydrological Administrative Regions: Region XI corresponds to the Rio Grijalva. Information is available not

only for the Rio Grijalva but also to the Usumacinta River, which joins the Rio Grijalva in the floodplain in Tabasco before it reaches the Gulf of Mexico.

Tabasco and Chiapas have the largest and second largest annual average precipitations in Mexico, 2406mm and 1969mm, respectively. The Grijalva-Usumacinta hydrologic region has by far the highest annual-average runoff in Mexico, 115,536 hm^3. (Papalopan follows, with only 44,662 hm^3). The area of watershed is estimated to be 128,390 km^2, the length of the river estimated as 1,911 km, and the altitude range from sea level to 4026m above mean sea level.

In general, the region has a humid, tropical weather in the lowland portion of the basin, with tall perennial rainforest whose area continues to decrease to make way for agriculture and ranching activities. However, the climate can also be temperate and foggy on the mountain ranges, allowing the development of cloud forests.

The climatic variability is high and associated with diverse meteorological phenomena, including the seasonal migration of the Inter-Tropical Convergence Zone (ITCZ), hurricanes, a mid-summer drought, easterly waves, and cold fronts (Magaña et al., 1999b;Uribe, 2002). Monthly precipitation (Fig. 1.1) is representative of southern and central Mexico. The rainy season starts in May, as the ITCZ approaches the coasts of Oaxaca and Chiapas (Uribe, 2002). There is subsequently a relative decrease in observed precipitation in July and August which characterizes the mid-summer drought (Magaña et al., 1999a). The rainy

season then ends in late October as the ITCZ moves southward. The region has been documented to be affected by the ENSO, which is associated with drier rainy seasons and wetter winters (Magaña et al., 1999b).

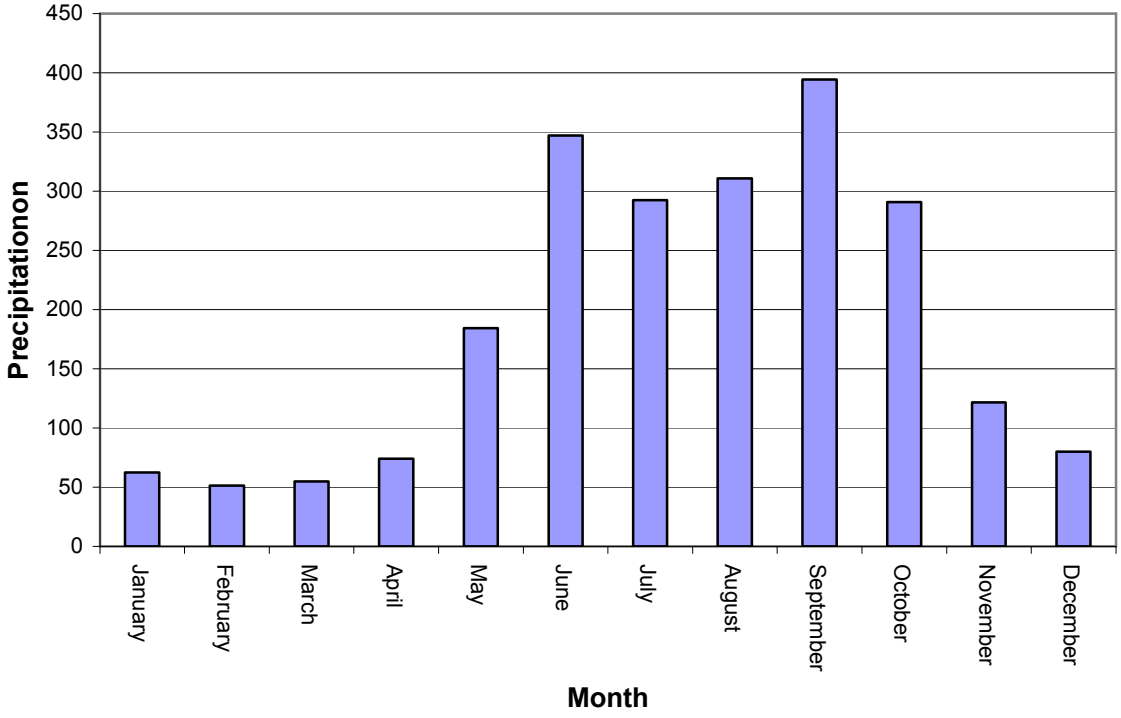
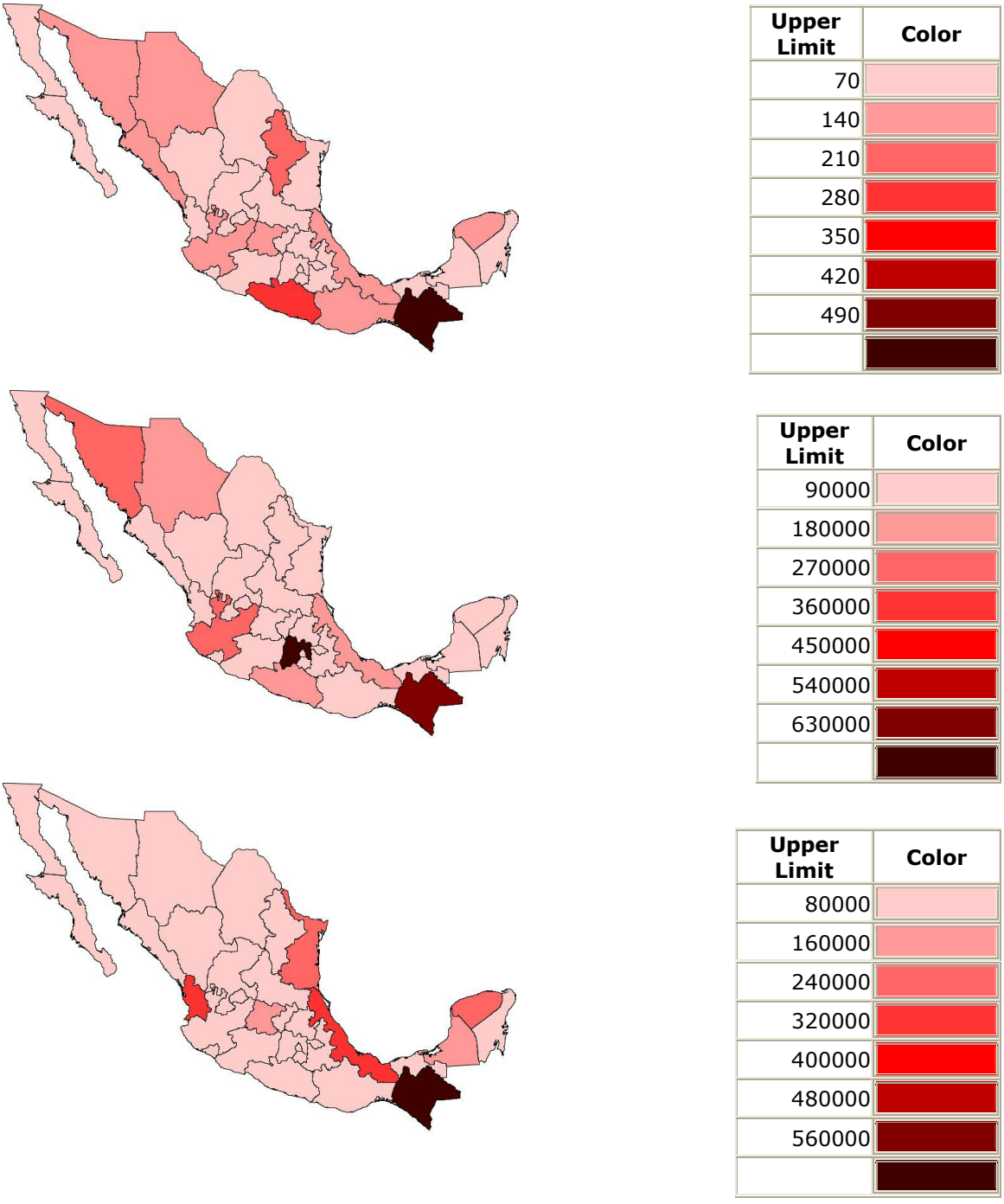


Figure 1.1. Average monthly precipitation [mm] in Hydrological Administrative Region XI (Grijalva-Usumacinta) from 1941-2005 (CONAGUA, 2006).

The heavy rainfall in this basin has both negative and positive aspects. The available documentation is once again limited and contradictory and, although not an exhaustive disaster inventory, the international inventory *Desinventar* (2007) was selected to estimate precipitation-related losses. Reports of individual events can be limited, so several events were grouped in an attempt

to get more representative results. In terms of deaths and hectares of damaged crops, Chiapas is consistently the Mexican state which is most affected by floods, flash floods, rainfall events, storms, hurricanes, hailstorms, and avalanches and the second state in terms of the people affected by these phenomena (Fig. 1.2). The impact of these events is enhanced by the proximity of the river to the two state capitals, Tuxtla Gutierrez and Villahermosa, which are the largest centers of urban development in the states of Chiapas and Tabasco, respectively.



Upper Limit	Color
70	
140	
210	
280	
350	
420	
490	

Upper Limit	Color
90000	
180000	
270000	
360000	
450000	
540000	
630000	

Upper Limit	Color
80000	
160000	
240000	
320000	
400000	
480000	
560000	

Figure 1.2. Numbers of deaths (upper plot), affected people (middle plot) and hectares of damaged crop (lower plot), caused by floods, flash floods, rainfall events, storms, hurricanes, hailstorms, and avalanches, between 1970 and 2000.

The construction of four dams was started in 1966, motivated by the large streamflow. The immediate purpose of the dams was to give hydropower production and to control floods. In decreasing order with altitude, these dams are usually referred to as La Angostura, Chicoasén, Malpaso, and Peñitas. Table 1.1 shows the individual power capacity and hydropower production for these four dams. The *Río Grijalva* basin is the most hydropower productive in the country and provides 52.30% of the national hydropower production.

Table 1.1. Dams with nicknames, power capacity and power for hydropower plants in the Rio Grijalva basin.

Dam (*Nickname*)	**Power Capacity [MW]**	**Power [kW h y^{-1}]**
Belisario Domínguez (*La Angostura*)	900	2025×10^{12}
Manuel Moreno Torres (*Chicoasén*)	1500	4500×10^{12}
Netzahualcóyotl (*Malpaso*)	1080	3000×10^{12}
Angel Albino Corzo (*Las Peñitas*)	420	1450×10^{12}

Although precipitation in the basin is high, the local population has only limited access to water because water supply services and sanitation are poor (Fig. 1.3), and domestic use and farming accounts for most of the water used in the region. This is a symptom of a more general hydrological paradox in Mexico. Hydrologic Administrative Regions in northern and central Mexico are the largest contributors (85%) to the Gross Domestic Product (GDP) but have limited water resources with only 32% of the national natural availability: southern regions, on the other hand, have more water resources, 68% of the national natural

availability, but are less developed and therefore make a smaller (15%) contribution to the GDP. Additional socio-economical aspects of the basin include the fact that it has tourism and ecological reserves. In fact, 12% of the territory of Chiapas has been declared an ecological reserve and the basin contains the highest documented biodiversity in Mexico and provides a habitat for 67% of the living species in the country.

All the aspects described above contribute to making this basin interesting from a hydrometeorological stand point. The large size of the basin implies it has a broad spectrum of hydrologic properties and is subject to a wide range of meteorological phenomena. Consequently, the application of distributed (rather than lumped) hydrological models, which are more interesting in the context of the present study, seems appropriate. However, the meteorological features that occur in this basin are similar to those which can occur elsewhere in the country. Consequently, if this study is successful here, there is possibility of similar success elsewhere in other Mexican river basins in the future. Within the basin itself, potential applications of the products derived from the successful creation of a hydrometeorological modeling system include providing improved warnings of floods and better water management, the latter being especially significant in the context of the basin's importance to Mexico's hydropower production network. For these reasons, this basin was selected as the preferred candidate for consideration in the present study.

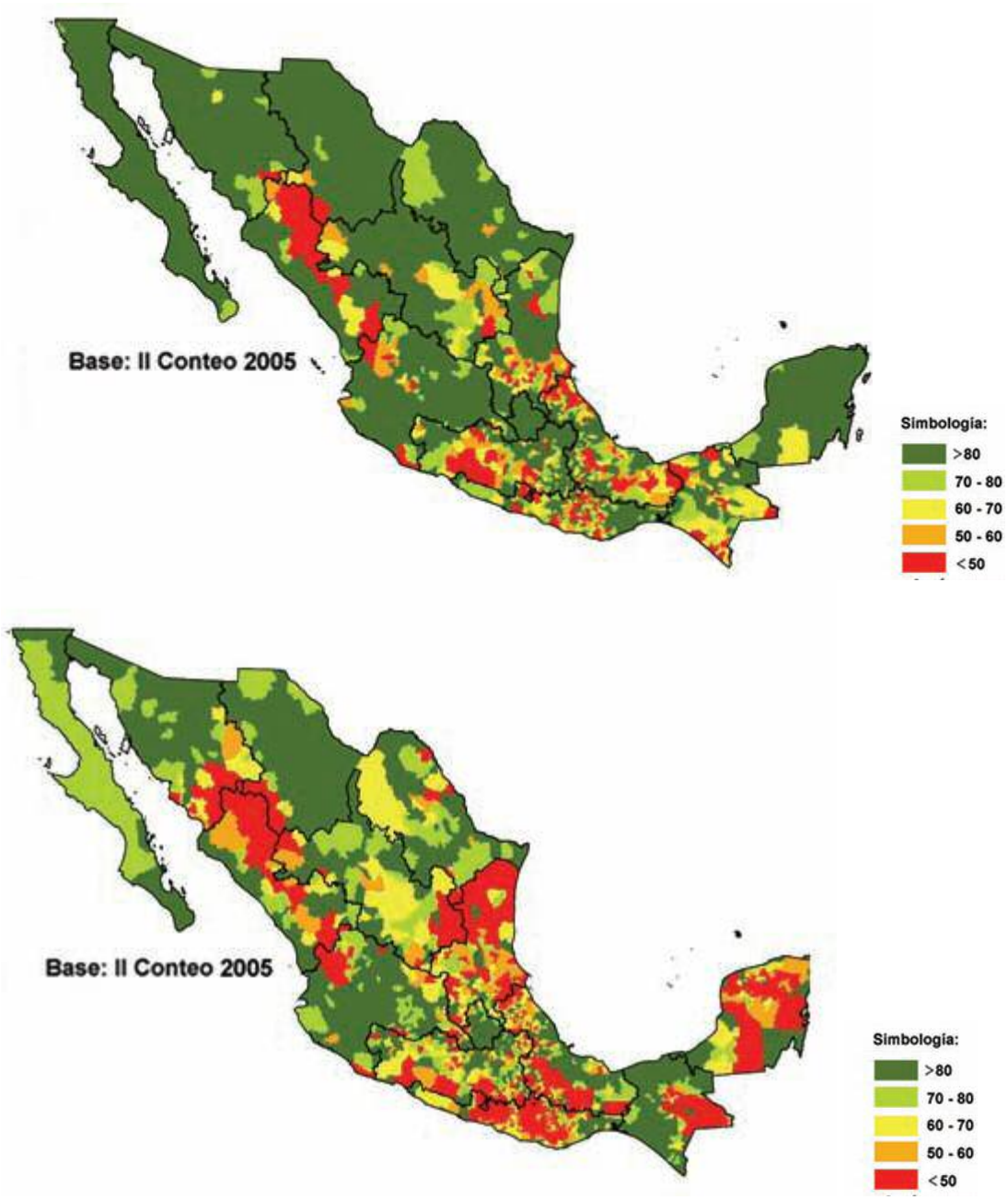


Figure 1.3. Perecentage of water supply (upper panel) and sanitation (lower panel) coverage per county in 2005 (taken from (CONAGUA, 2006)). The black line indicates division between Hydrologic Administrative Regions.

1.3. Objective and General Approach

The objective of this study was to investigate establishing a short- to medium-term hydrometeorological forecasting system for the *Río Grijalva* basin. Daily modeled meteorological fields from the North American Regional Reanalysis (NARR) were implemented to drive a hydrological model which was used to estimate streamflow. Using these NARR fields as forcing variables provides a convenient substitute for the forcing fields that would otherwise have to be calculated by a regional Numerical Weather Prediction Model (NWPM). Their use removed the need to develop such a regional NWPM and allowed more time and resources to be focused on the development of the (arguably more problematic) hydrologic component of the proposed hydrometeorological forecasting system. In addition, the long data record provided by NARR (3-hr fields starting in 1979) facilitates the validation of the methods proposed for (a) establishing the hydrologic model and (b) implementing the meteorological fields (it is possible to exploit the better documentation of NARR fields relative to those which would otherwise need to be derived from a regional NWPM).

Establishing a hydrological model necessarily starts with the selection of the model, which is most appropriately made on the basis of the hydrological processes of interest. However, in poorly documented basins, data availability becomes an important factor. The hydrologic model selected for the experiments is MMS-PRMS. This model was selected as a physically-based, distributed hydrological model that is highly accessible and customizable, with valuable

features such as modular constitution, computing-platform versatility, time-step versatility, and public-domain documentation and code. In addition, MMS-PRMS comes with both a parameterization tool (GIS-Weasel) and a parameter optimization tool (LUCA) based on the highly effective Shuffled Complex Evolution Method (SCE), both of these being essential tools for establishing and improving the model.

The initial step in this study was to propose and create an alternative parameterization method for MMS-PRMS, given the fact that the standard datasets used by GIS-Weasel are only available in the US. The initial model thus obtained must be forced with meteorological data; and an adequate method for using the intermittent reports from scattered meteorological stations was necessary in this basin. The model representation was then improved through parameter optimization using the LUCA tool. This initial portion of the study resulted in a calibrated hydrological model for the *Río Grijalva* basin.

The second potion of the study involved implementing the NARR-fields as forcing for this hydrological model. Since observed streamflow is not available for the period when NARR data are available, it was necessary to compare modeled streamflow when using NARR data and when using data from local meteorological stations. Similar implementation methods for both datasets were required so that any differences in modeled streamflow could be associated solely with biases in NARR data. A bias analysis was then performed to define the essential characteristics of a calibration methodology for NARR-rainfall

estimates prior to their use in the hydrological model. Finally, because using the NARR data does not provide ensembles modeled fields, a method for the estimation of uncertainty in modeled streamflow based on the bias analysis was explored.

The methods proposed and developed in this should be exportable to other river basins in Mexico basin (and hopefully other international basins), and they provide the basis of a future hydrometeorological forecasting system for the *Río Grijalva* basin in which the NARR fields would be substituted by fields calculated by a regional Numerical Weather Prediction Model.

2. MODEL COMPONENTS

2.1. Hydrological Modeling

2.1.1. The Modular Modeling System (MMS)

The development of the MMS by the USGS and the University of Colorado started in 1989 (Leavesley et al., 1996). A large number of agencies, such as the U.S. Bureau of Reclamation, the U.S. Forest Service, the National Aeronautics and Space Administration, the Agricultural Research Service, and the Terrestrial Ecology Regional Research and Analysis (TERRA) Laboratory, subsequently joined the effort. The objective was to create a framework for model development by: (1) addressing the generation, testing and evaluation of algorithms that represent physical process (especially processes related to energy, water, chemistry, and biology); and (2) facilitating the integration of customized algorithms into operational models.

2.1.2. The Precipitation Runoff Modeling System (PRMS)

PRMS is a precipitation-runoff model built under MMS. PRMS is a deterministic, physically-based, distributed model developed to evaluate the impacts of meteorological conditions and land-use on the hydrologic components of a basin, such as streamflow, sediment yield, water balance relationships, flow regimes, flood peaks and volumes, soil-water relationships, and groundwater recharge (Leavesley et al., 1983). PRMS is highly customizable because it consists of modules addressing individual processes within the hydrological

cycle. The model subdivides the basin into Hydrologic-Response Units (HRUs) on the basis of physical characteristics such as, slope, aspect, soil, and vegetation. Water and energy balances are performed daily for each HRU, and the sum of all the area-weighted responses corresponds to the response of the watershed.

There are different versions of the PRMS and the two following sections relate to the particular version of PRMS used for the present study. The model used is a daily version of the PRMS in which some of the components were disabled including the modules that related to snow (snow is negligible in the Rio Grijalva Basin) and rainfall interpolation (a customized approach was developed for the interpolation of the data in this study, see the Methodology section).

2.1.2.1. Conceptual Model

Figure 2.1 is a diagram of the conceptual model which underlies PRMS. Input variables include daily maximum and minimum temperatures, and daily total precipitation. The model consists of interconnected reservoirs that represent the different components of the hydrologic system (vegetation, soil, impermeable surfaces, etc.).

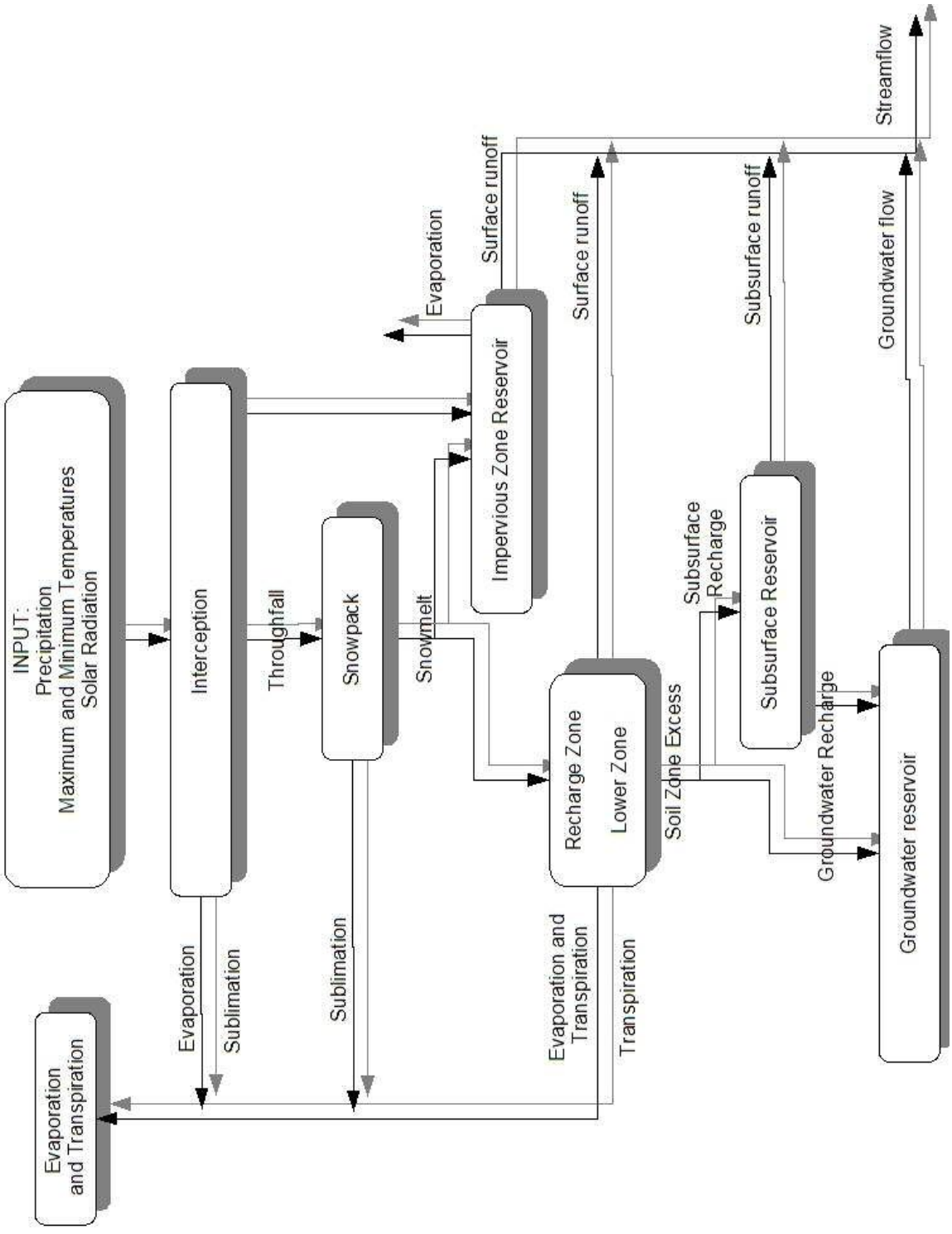


Figure 2.1. Schematic diagram of the conceptual watershed system behind MMS-PRMS (Leavesley et al., 1983).

Precipitation is the most important driving variable. An initial portion of the precipitation is intercepted by vegetation and the remainder falls on the surface. A portion of the surface can be defined as an impermeable reservoir which produces surface runoff immediately when its capacity is exceeded. The amount retained by both the impermeable reservoir and the vegetation reservoir can be depleted by evaporation.

The net amount of rainfall received by the soil reservoir is partitioned into infiltration and runoff. Water infiltrated is stored but can be lost due to evaporation (upper and lower layer) and transpiration (upper layer only). Surface runoff is estimated using a contributing-area concept, and infiltration is estimated as the difference between net rainfall and runoff. The maximum volume that can be stored by the soil reservoir is equivalent to the field capacity, and the minimum, to the wilting point. Excess water in the soil layers is partitioned between the subsurface and groundwater reservoirs. Water received by the subsurface reservoir is further partitioned into percolation to the groundwater reservoir and to subsurface discharge downstream. The discharge is considered to be water in the saturated-unsaturated zones moving relatively fast to a channel system.

As mentioned above, recharge to the ground-water reservoir occurs from both the soil zone and the subsurface reservoir. There is a daily upper limit on the amount of water transferred from the soil to the groundwater reservoir. This limit is a function of a recharge rate coefficient and the volume of water stored in

the subsurface reservoir. The ground-water reservoir is the source of all baseflow. Finally, a sink in the ground-water reservoir is defined to handle the movement of water beyond the area of interest.

2.1.2.2. Mathematical Model

The present section describes the processes modeled by each of the modules in PRMS. Below Fortran files are indicated in italics by the suffix "*f* " after the name of the module and parameters are indicated in italics.

Interception (*intcp_prms.f*): Depends on the seasonal vegetation cover density (parameters *covden_sum* and *covden_win* for the summer and winter respectively) and the seasonal vegetation storage (*srain_intcp* and *wrain_intcp*). Summer net rain (*net_*rain), for example, is defined as:

$$net\_rain = \left[hru\_rain \cdot \left(1 - covden\_sum\right)\right] + \left(thrufall \times covden\_sum\right),$$

where *hru_*rain is the average depth of rainfall over a HRU, and thrufall is derived from:

$$\begin{aligned} thrufall &= hru\_rain - (srain\_intcp - intcp\_stor) && \text{for} \quad hru\_rain > (srain\_intcp - intcp\_stor) \\ thrufall &= 0 && \text{for} \quad hru\_rain \le (srain\_intcp - intcp\_stor) \end{aligned},$$

where *intcp_stor* is the amount of water being stored in the canopy. The rate at which the intercepted water evaporates is proportional to the estimated potential evaporation (*potet*; see evaporation section below) divided by a monthly evaporation-pan coefficient (*epan_coef*):

$$evcan = \frac{potet}{epan\_coef}.$$

Potential Evapotranspiration (*potet_hamon_prms.f*): A period is specified where evapotranspiration (ET) is expected to occur every year. The maximum temperatures are accumulated daily starting the first day of this period. When the cumulative value exceeds a certain threshold value (*transp_tmax*), transpiration starts. The potential ET is estimated from the daily mean air temperature (*tavgc*) and possible hours of sunshine (*radpl_sunhrs*) through the following equations.

$$potet = hamon\_coef \cdot radpl\_sunhrs^2 \cdot vdsat,$$

where *hamon_coef* is a monthly coefficient, and *vdsat* represents the saturated vapor density, which can be estimated from

$$vdsat = 216.7 \cdot \frac{vpsat}{tavgc + 273.3}.$$

Saturated vapor pressure (*vpsat*) represents the saturated vapor pressure, estimated from

$$vpsat = 6.108 \cdot \exp\left[17.26939 \cdot \frac{tavgc}{tavgc + 237.3}\right].$$

Surface Runoff (*srunoff_smidx_prms.f*): Runoff is estimated using a non-linear contributing area (*ca_percent*) concept through the following equation

$$ca\_percent = smidx\_coef \cdot 10^{smidx\_exp \cdot smidx},$$

where *smidx_coef* is a non-linear coefficient, *smidx_exp* is a exponent, and *smidx* is defined as

$$smidx = soil\_moist \cdot \frac{net\_rain}{2}.$$

Parameter *soil_moist* represents soil moisture in the previous equation. Surface runoff from the pervious area (*srp*) is computed as

$$srp = ca\_percent \cdot net\_rain,$$

where *net_rain* symbolizes net rainfall. Infiltration is defined as net rainfall minus surface runoff

$$infil = net\_rain - srp.$$

Soil Moisture Accounting (*smbal_prms.f*): Soil Moisture is the summation of water accretions (net rainfall, throughfall) and depletions (ET, and subsurface and groundwater recharges) from the soil profile. The maximum water depth stored in this reservoir (*soil_moist_max*) is equal to the difference between field capacity and wilting point of the soil. The soil is divided into an upper layer (or recharge zone) and a lower layer. The water depth capacity of the upper layer is *soil_rechr_max*. The difference between *soil_moist_max* and *soil_rechr_max* gives the water holding capacity of the lower zone. After infiltration is added to the soil zone, water exceeding *soil_moist_max* is distributed between the subsurface and groundwater reservoirs. Exceeding water is used to satisfy a maximum daily recharge of the groundwater reservoir (*soil_to_gw*), the remaining

portion if any is added to the subsurface reservoir. Actual evapotranspiration (*hru_actet*) reflects the availability of water to satisfy potential ET. The order in which water is taken to satisfy the demand is the following: (1) Evaporation from water intercepted by the vegetation, (2) evaporation from impervious storages, and (3) soil storage. ET from the soil layers is calculated from theoretical relationships between water available in the soil layer (either *soil_moist* or *soil_rechr*) and the ratio between actual and potential ET for different soil types (sand, loam, clay).

Groundwater Flow (*gwflow_prms.f*): Groundwater is conceptualized as a linear reservoir responsible for baseflow. Inflow to this reservoir is generated when infiltration exceeds the water holding capacity of the soil layers, and when water is available in the subsurface reservoir. The flow from the groundwater reservoir (*gwres_flow*) is estimated from:

$$gwres\_flow = gwflow\_coef \cdot gwres\_stor ,$$

where *gwflow_coef* represents a coefficient; and *gwres_stor*, the total storage in each groundwater reservoir. Groundwater moving beyond the area of interest is treated as a sink computed by:

$$gwres\_sink = gwsink\_coef \times gwres\_stor ,$$

where *gwsink_coef* is a coefficient.

Subsurface Flow (*ssflow_prms.f*): The flow is estimated through a reservoir routing system. The continuity of mass for the subsurface flow is expressed as:

$$ssres\_flow = ssres\_in - \frac{d\left(ssres\_stor\right)}{dt},$$

where *ssres_flow* represents the contribution to streamflow from the subsurface reservoir, *ssres_in* is the total inflow to the reservoir, and *ssres_stor* is the amount of water stored by the reservoir. On the other hand, the following empirical relationship can be used to estimate (*ssres_flow*)

$$ssres\_flow = \left(ssrcoef\_lin \cdot ssres\_stor\right) + \left(ssrcoef\_sq \cdot ssres\_stor^2\right).$$

The last two equations can be combined to solve for *ssres_flow*. The discharge from the reservoir to the groundwater reservoir (*ssr_to_gw*) is assumed to follow the relationship:

$$ssr\_to\_gw = ssr2gw\_rate \cdot \left[\frac{ssres\_stor}{ssr2gw\_\max}\right]^{ssr2gw\_exp},$$

where *ssr2gw_rate* and *ssr2gw_exp* are coefficients, and *ssr2gw_max* is the maximum value for water routed from subsurface to groundwater. The parameter *ssr2gw_exp* can be set equal to one to make the function linear.

Streamflow (*strmflow_st_prms.f*): Basin streamflow (*basin_cfs*) is computed from

$$basin\_cfs = basin\_sroff + basin\_ssflow + basin\_gwflow,$$

where *basin_sroff*, *basin_ssflow*, and *basin_gwflow*, are area-weighted averages of surface, subsurface and groundwater flows respectively. Channel-flow routing is performed only when surface reservoirs are designated. Reservoir routing is based on the continuity equation:

$$O = I - \frac{d\left(sfres\_sto\right)}{dt},$$

where *O*, *I* and *sfres_sto* represent respectively the outflow, input and storage of the reservoir. The input (*I*) is defined as the sum of the streamflow contributions from all the HRUs and the outflows from both the subsurface and groundwater reservoirs upstream. Inflow can include the outflow from up to three upstream channel reservoirs. A linear storage routing defined from the following equation is used:

$$sfres\_outq = sfres\_coef \cdot sfres\_sto,$$

where *sfres_coef* represents a coefficient. The two previous equations can be combined to solve for *sfres_outq*.

2.1.3. The GIS-Weasel for Basin Parameterization

Weasel consists of algorithms in Arc Macro Language (AML), awk, C and Unix, interacting with Arc/Info to provide a Graphic User Interface (GUI) for the delineation, characterization, modification, and parameterization of a basin (Viger et al., 1998). The two critical components from Weasel used in the present work include only basin delineation and parameterization of the basin.

For the delineation of the basin, a Digital Elevation Model (DEM) is necessary. The first processing step performed by Weasel consists of filling depressions in the DEM to ensure all the flow paths will lead to the edge of the elevation model. The resulting model is used to create the flow direction and accumulation surfaces, which indicate the direction of outflow from each cell, and the number of upslope contributing cells respectively.

The watershed is derived from the previous surfaces and a user-specified pour point, which usually corresponds to a streamflow station targeted for simulation. The watershed can also be defined by using a pre-existing geo-dataset.

The drainage network can be derived from the accumulation surface and a user-specified threshold area that represents the minimum area upslope to initiate a first-order link segment of the drainage network. The links are defined as segments of the drainage network that begin and end with a node. Nodes in the other hand occur at 1) the points of first order links, 2) the concurrence of two or more links, and 3) the basin outlet. Lower (higher) area thresholds will give higher (lower) density drainage networks.

The parameterization component of Weasel is able to derive parameters related to all the spatial features of the basin like: the Area of Interest (AOI), the Hydrologic Response Units (HRUs), segments and nodes of the drainage network, hydrometeorological stations, wells, user-defined grids, and their combinations.

Specialized routines are available for the derivation of PRMS parameters related to topography, geography, soil and vegetation. Parameters related to geography and topography can be derived from regular DEMs. The subroutines for the derivation of soil and vegetation parameters, on the other hand, support datasets only available for the United States, like the STATSGO (State Soil Geographic Database), SSURGO (Soil Survey Geographic Database), and the National Land Cover Characterization. GIS-Weasel can still derive soil and vegetation parameters from surrogate datasets but some specific parameters must be provided beforehand. The specific parameters include: land cover reclassifications, root depth, seasonal vegetation, interception storages, seasonal vegetation cover densities, maximum soil moisture, storage capacity of the upper layer, and dominant soil texture.

2.1.3.1. Surrogate Datasets for the GIS-Weasel

As mentioned previously, MMS-PRMS requires topography, soil, and vegetation parameters. However, the standard datasets for the parameterization are only available for the United States. Thus the surrogate datasets described below are used for the parameterization of the Rio Grijalva basin:

Topographic Parameters: The Shuttle Radar Topography Mission (SRTM (2006b;Rodriguez et al., 2005)) dataset is used for the derivation of the topographic parameters. The SRTM is a project developed by both the National Geospatial-Intelligence Agency (NGA) and the National Aeronautics and Space

Administration (NASA). The objective of SRTM is to produce digital topographic data for land areas between 60° north and 56° south latitude (~80% of the land surface of the Earth). The target grid size is 1-arc second (approximately 30 meters) on a latitude/longitude grid. The dataset used in the present work, however, is the 3-arc second (approximately 90 meters), which is the only data available outside the US territory through the *Seamless Data Distribution System* (http://seamless.usgs.gov/website/seamless/index.asp).

Vegetation Parameters: The Global Land Cover Characterization Data Base (GLCC) is used in the present work. The GLCC is derived from the 1-km Advanced Very High Resolution Radiometer (AVHRR) by the U.S. Geological Survey's (USGS) National Center for Earth Resources Observation and Science (EROS), the University of Nebraska-Lincoln (UNL) and the Joint Research Centre of the European Commission (2006a). The AVHRR spans a 12-month period (April 1992 - March 1993). The corresponding monthly Normalized Difference Vegetation Indexes (NDVIs) are also available for the public. The datasets are available in two geographic projections: 1) Interrupted Goode Homolosine, and 2) Lambert Azimuthal Equal Area.

Soil Parameters: A soil properties database derived from the FAO (Food and Agriculture Organization) Digital Soil Map of the World (SMW) using global soil profile databases and pedo-transfer functions was used in this study (Reynolds C.A. et al., 1999). The database contains physical properties for two soil layers of depths (0-30 and 30-100 cm). The properties include: available

water content, particle-size distribution, dominant soil texture, organic carbon content, coarse fragments, bulk density, and porosity. The database is distributed in grids with pixel-size equal to five minute to preserve the size of the SMW.

2.1.4. The "Let Us Calibrate" (LUCA) Tool for Parameter Optimization

LUCA is a GUI (Graphical User Interface) for step-wise, multi-objective, parameter optimization tool for the MMS-PRMS in development by the USGS (Hay et al., 2006b;Hay et al., 2006a). The algorithm is based on the Shuffled Complex Evolution Method (SCE) developed at the University of Arizona (Duan et al., 1992;Duan et al., 1994;Duan et al., 2006). The version of LUCA used in the present work was provided by Lauren Hay from the USGS (Hay, 2006).

LUCA provides several means for highly customizable optimization strategies. Different combination of parameter sets can be calibrated during different steps, and the same number of steps can be performed several times (rounds). Distributed parameters can be calibrated either individually or collectively. The collective optimization is performed by exploring the mean value of the parameter and distributing the difference to the original value proportionally among the parameters. There are four Objective Functions (OFs) currently implemented: Normalized Root Mean Square Error (NRMSE), Nash-Sutcliffe, Partially Normalized Absolute Difference (PNAD), and the logarithmic absolute difference. The inclusion of different OFs also adds to the strategy capabilities

because different OFs are expected to perform better for different regions of hydrographs. Additionally, the OFs can be combined for multi-objective parameterization. Elements of the SCE, including number of complexes, simplex and complex sizes, as well as termination criteria, can also be defined by the user. Periods and time steps for the OF evaluation are also customizable.

2.2. Hydrometeorological Data

2.2.1. The Local Meteorological Network

Two meteorological station networks are used in the present work: *Extractor Rápido de Información Climática* (Fast Extractor of Climatic Information, ERIC) from Mexico and a database of the *Instituto Nacional de Sismología, Vulcanología, Meteorología e Hidrología* (National Institute for Seismology, Volcanology, Meteorology, and Hydrology, INSIVUMEH) from Guatemala.

ERIC is provided in a CD-ROM that contains both the climatic information and the software to access it. ERIC reports daily data recorded at meteorological stations throughout Mexico between 1930 and 1990. The CD-ROM is edited by the *Instituto Mexicano de Tecnología del Agua* (Mexican Institute for Technology of Water, IMTA (Quintas, 1996)). Reported variables include maximum and minimum temperatures, rainfall, evaporation, and cloudiness.

ERIC only covers Mexican territory so it needs to be complemented with Guatemalan data from the INSIVUMEH that reports daily maximum and minimum temperatures, and rainfall.

2.2.2. Surface Hydrology Database

Streamflow measurements and volume stored at dams are extracted from the *Sistema de Información de Aguas Superficiales* (Surface Water information System, SIAS also known as *Banco Nacional de Datos de Aguas Superficiales*

(National Bank of Surface Water Data, BANDAS)) edited by IMTA (1998). BANDAS is distributed in a CD-ROM. It contains information on streamflow stations and volumes at dams throughout Mexico. Information on localization, installation date, measurement methodology, and other relevant comments on the station are reported as well.

2.2.3. North American Regional Reanalysis (NARR)

The NARR has been defined as a "long-term, dynamically consistent, high-resolution, high-frequency, atmospheric and land surface hydrology dataset for the North American domain" (Mesinger et al., 2006). The objective of NARR is to provide information about the variability of water, especially precipitation, in the U.S. NARR is expected to capture extreme events, and to interface well with hydrological models.

Temporal coverage from NARR starts in 1979. The grid system has a pixel size of 32km, with vertical resolution of 45 layers, and data are available every three hours. The critical additions to NARR are the lateral boundary conditions and data from the NCEP/DOE Global Reanalysis, the NCEP Eta Model (and its Data Assimilation System), a recent version of the Noah LSM, and additional or improved data sets not contained in the Global Reanalysis. The most important addition to NARR is, however, the assimilation of observed precipitation. Observed precipitation is assimilated as latent-heat in an attempt to

ensure modeled hydrological variables are more realistic than if the model was free to forecast precipitation (Lin et al., 1999).

The 24-hr analysis of precipitation from rain gauges is disaggregated into hourly analysis before assimilation. Two different analysis and dissagregation techniques are employed. Over Mexico and Canada the 1 degree analysis is generated using the Cressman successive-scan analysis technique. This analysis is disaggregated using the GR2 1-hourly precipitation forecasts. Over the continental United States (CONUS), the 1/8-th degree analysis is obtained using an inverse square-distance weighting scheme and an orographic enhancement technique known as the Parameter-elevation Regressions on Independent Slopes Model (PRISM (Daly et al., 1994)). This analysis is disaggregated using temporal weights derived from a 2.5-degree analysis of hourly rain gauge data.

Note worthy features of NARR relative to the NCEP-DOE Global Reanalysis (GR2) include a more realistic representation of the precipitation over CONUS, fits of tropospheric temperatures and winds to rawinsonde observations throughout the troposphere, and fits of 2m temperatures and 10m winds to surface station observations. NARR developers suggest the most important factors responsible for better results are in order of importance (1) more information being assimilated; (2) better assimilation techniques; (3) enhanced resolution; and (4) a better model.

3. METHODOLOGY

Establishing the hydrological component of the proposed hydrometeorological system involves three steps: (1) parameterizing the hydrological model for the selected basin, (2) acquiring available data from meteorological stations, (3) interfacing the meteorological fields into the hydrological model, and (4) optimizing parameters in the hydrological model.

3.1. Basin Parameterization

Basin parameterization involves processing of grids with different projections and cell sizes. The GIS-Weasel manual suggests using a geographic projection that ensures the sides of the cells are approximately equal throughout the grid (resulting in nearly square areas). The Cylindrical Equal Area - Geographic Projection (CEA-GP) ensures no area distortion and equally spaced meridians by having parallels farthest apart near the equator. However, the variation in the spacing of the parallels is very small across the Rio Grijalva basin, and the cell sides are similar also. Thus the CEA was selected to be the common geographic projection used for all the geo-datasets used in this project. When grids with different cell sizes were combined during the parameterization process, the cell size of the highest resolution grid to be combined was adopted and used as the cell size of the combined grid.

Derivation of the topographic parameters from the SRTM was made using the GIS-Weasel after CEA projection. Because the standard datasets for the

derivation of soil and vegetation parameters normally provided to GIS-Weasal are not available for Mexico, parameter grids were derived from the surrogate datasets described in Chapter 2 prior to their use by the GIS-Weasel to (1) estimate parameter values for each HRU, and (2) create the input parameter file required by MMS-PRMS. The vegetation parameters required by MMS-PRMS include land cover reclassifications, root depth, seasonal vegetation, interception storages, and seasonal vegetation cover densities. Although GIS-Weasel is not able to recognize the surrogate datasets directly, the GLCC land cover classification is in fact very similar to the classification in the standard datasets that Weasel can use. Therefore, it was possible to follow the parameter derivation performed by GIS-Weasel closely. Hence land cover reclassifications, root depth, seasonal vegetation, and interception storages were derived from the GLCC using the lookup tables originally established for the GIS-Weasel that relate land cover type to the parameter values (http://wwwbrr.cr.usgs.gov/weasel/manual/appendix_a.html).

The derivation of Seasonal Vegetation Cover Densities (SVCDs) is not straightforward. GIS-Weasel estimates *SVCD* from the following formula:

$$SVCD = a \cdot VCD ,$$

where *VCD* is the vegetation cover density, and *a* is a weight that can be obtained from the lookup tables established for GIS-Weasel. VCD is estimated from the median of the monthlies NDVI grids. It is assumed that the highest monthly NDVI cell corresponds to regions with 100% vegetation density cover

while the 0% vegetation cover density is assigned to cells with the lowest values. Hence the VCD grid can be estimated from:

$$[VCD] = \frac{\left[\overline{NDVI}\right] - NDVI_{\min}}{NDVI_{\max} - NDVI_{\min}},$$

where $NDVI_{min}$ and $NDVI_{max}$ are the minimum and maximum values found in the median NDVI grid $\left[\overline{NDVI}\right]$. In the Río Grijalva basin, it is expected that $NDVI_{min}$ will correspond to inland water bodies while $NDVI_{max}$ will correspond to areas of rainforest.

The soil parameters required include the maximum soil moisture (*soil_moist_max*), the storage capacity of the upper layer (*soil_rechr_max*), and dominant soil texture. Maximum soil moisture is defined in MMS-PRMS as the storage depth of water for the layer of soil with depth equal to the rooting depth. On the other hand, available water holding capacity from the Reynolds dataset is defined as the depth of water in the upper meter of the soil layer. If the AWHC is normalized per unit length, parameter *soil_moist_max* can be estimated from:

$$soil\_moist\_max = AWHC \cdot root\_depth,$$

where *rooth_depth* is the root depth previously derived. The parameter *soil_rechr_max* is estimated from:

$$soil\_rechr\_max = AWHC \cdot recharge\_root\_depth,$$

where *recharge_root_depth* is either the rooting depth or 18in, whichever is least, this being consistent with the rule used in GIS-Weasel.

Finally, GIS-Weasel defines the dominant soil texture as clay, if the portion of clay is greater than 40%; as sand, if the portion is greater than 50%; and silt otherwise. In this study, the dominant soil texture was derived from the fraction of clay, silt and sand taken from the Reynolds dataset following the same rules. However, since the Reynolds dataset reports the fractions for two upper soil layers (0-30cm and 30-100cm), the 0-100cm weighted mean was used to determine the dominant texture.

3.2. Acquiring Meteorological Data

MMS-PRMS requires input files in ASCII format with the meteorological data associated with HRUs aligned in columns. However, providing data which follows this convention is operationally difficult in the Rio Grijalva basin because there are a large numbers of stations and they often provide intermittent data. To provide a consistent dataset, the ERIC meteorological stations were interpolated on to regular grids. Uribe (Uribe, 2000) showed that such interpolated fields of rainfall and temperature from ERIC generated using kriging reproduce relevant climatic features in Mexico, such as the inter-annual variability, the Mid-Summer Drought (Magaña et al., 1999a), and the effect of both topography and ocean-continents boundaries. Furthermore, the resulting regular grids are more suitable for modeling purposes and the average value for each HRU required by MMS-PRMS can easily be estimated from the interpolated grid by averaging the cells that correspond to each HRU.

The procedure to generation of regular gridded data started with the generation of daily files which list the location and measurements of the active meteorological stations. The interpolation was performed through the *kriging* command built into Arc/INFO using default options, except for the cell size required for the resulting grid. The cell size selected was defined as a multiple of the HRU grid so that estimating the average values for each HRU would not require further interpolation. Since both the HRUs-grid and the interpolated grid have similar projections and cell sizes, the HRUs can be easily mapped onto the interpolated grid and the values for appropriate cells averaged to provide the meteorological variables for each HRU.

3.3. Interfacing Modeled-Estimated Meteorological Fields to MMS-PRMS

Although model-estimated meteorological fields (e.g. those from NARR) are usually reported on a regular grid, the data still need to be processed before they can be applied with the MMS-PRMS because of (1) the large number of points associated with a single HRU, (2) the need to map the NARR grids onto the HRUs grid, and (3) the need to project the source data onto CEA-GP. In practice all of these issues can be easily solved by treating every cell node as a meteorological station and then applying the procedure used for meteorological stations described in the previous section. The benefit of doing this is two-fold: first, there is no need to develop new algorithms and, more importantly, second,

by following the same procedure, data interfacing can no longer be a potential source of difference when the subsequently modeled streamflows are compared.

3.4. Parameter Optimization

When optimizing the MMS-PRMS, several challenges must be faced. One important challenge lies in the number of parameters involved (23 parameters for each HRU). Every parameter adds a dimension to the parameter space that must be searched. A second challenge lies in the relative uncertainty in the initial parameter set because of its derivation from surrogate datasets rather than the standard datasets normally used for the GIS-Weasel. An additional challenge is that optimization with LUCA of all model parameters in a single step tends to result in the selection on unrealistic values, suggesting the parameter search will be aided by using a partition strategy. Although several optimization strategies, based on sensitivity analysis and other criteria frequently applied in optimization, were attempted, the following strategy resulted in both realistic parameter values and the largest improvement of the selected OF (see definition of the PNAD below).

It was decided to partition the optimization process into “steps” and “rounds”, with the parameters in individual model modules calibrated during a succession of different steps that make up a round, and with optimization of the mean of distributed parameters at the first round and distributed individual parameters at a second round. Two important aspects of a partition optimization

strategy are: (1) how the partition is made, and (2) the order in which the different partitions are calibrated. The most obvious way to partition is in terms of model modules. It is assumed correlation between parameters in the same model module is high, while correlation between parameters in different model modules is low, consequently modules can be optimized individually in consecutive steps. The exceptional module is that related to soil moisture, which is continuously interacting with all the other modules so its parameters are expected to be correlated with parameters in other modules. It was therefore decided to include the soil moisture parameters in every optimization step.

The second important aspect of a partition optimization procedure is the order in which modules are optimized. It was decided it was best to start with the groundwater modules and finish with the modules related to evapotranspiration, thus reflecting an upward movement through the system (the so-called “bottom-up” optimization procedure). The reasoning behind this strategy is that the hydrograph is composed of several frequencies which can be associated with different modules. The groundwater module, for example, is responsible for the low-frequency component of the hydrograph (baseflow). It is expected that the improvement in modules associated with higher frequency variations in the hydrograph will be more effective once the modules associated with lower frequency variations have been improved through parameter optimization.

As mentioned above, rounds of optimization were used. The first round is meant to optimize the mean values of the distributed parameters, while the

second round is meant to improve individual distributed parameters (a “top-down” optimization procedure). The first round provides a general improvement in the parameters, while the second provides a detailed improvement. Once again, it is assumed a general improvement of the parameters will provide the basis for a more effective optimization of individual parameters. In summary, the strategy used for model optimization in this study comprised an individualized “bottom-up”, “top-down” optimization procedure, which was adopted with the objective of breaking-down the optimization problem into smaller, hopefully solvable, optimization problems.

In addition to allowing rounds and steps, LUCA also allows customizable optimization around four different Objective Functions (OFs) (see section 2.1.4). An important criteria when selecting an OF is the aspect of the hydrograph for which improvement is sought (Gupta et al., 1998). However, all the OFs provided by LUCA are expected to enhance deficiencies in the baseflow component of the modeled hydrograph. Consequently, the selection is relatively constrained and the OF selected in this study was the Partially Normalized Absolute Difference (PNAD) defined as:

$$PNAD = \sum_{i=1}^{n} \frac{\left| O_i^{obs} - O_i^{mod} \right|}{O_i^{obs}},$$

where O_i^{mod} and O_i^{obs} are the daily modeled and observed streamflow, respectively, and *n* is the number of daily observations.

3.5. Modeling Timeline

For better clarity, it is important to describe the periods during which modeling was carried out in the present study. The periods selected for modeling are primarily determined by data availability, but other factors are also involved. Figure 3.1 illustrates the modeling timeline which is discussed in detail below.

In this study, the first step was to set up the hydrological model, which involved model spin-up for initialization, parameter optimization, and model evaluation. The datasets involved in doing this are ERIC (daily meteorological observations used as forcing) and SIAS (daily streamflow used as a benchmark). ERIC reports from 1930 to 1990, and SIAS from 1948 to 1999. Although the data availability suggests modeling is possible from 1948 to 1990, the construction of dams fundamentally altered measurements at the target streamflow station after 1966. Thus, it is only possible to use the datasets from 1948 to 1965 and, for the purpose of model set up, this period was further broken-down into three sub-periods during which the data were used for initialization (1948), parameter optimization (1949-1952), and evaluation (1979-1965).

The second step involved using the modeled-estimated meteorological fields from the NARR to force the hydrological model. NARR data are available from 1979 to the present day. Unfortunately, streamflow measurements from SIAS are not available for this period: the only dataset available for comparison is ERIC. For this reason, it was assumed that the streamflow calculated using ERIC data by the model (set up using ERIC and SIAS data from 1948-1965) could be

used as a benchmark against which to judge the modeled streamflow calculated using data from NARR applied to the same model. The validity of this assumption clearly depends on the reliability of the optimized hydrological model.

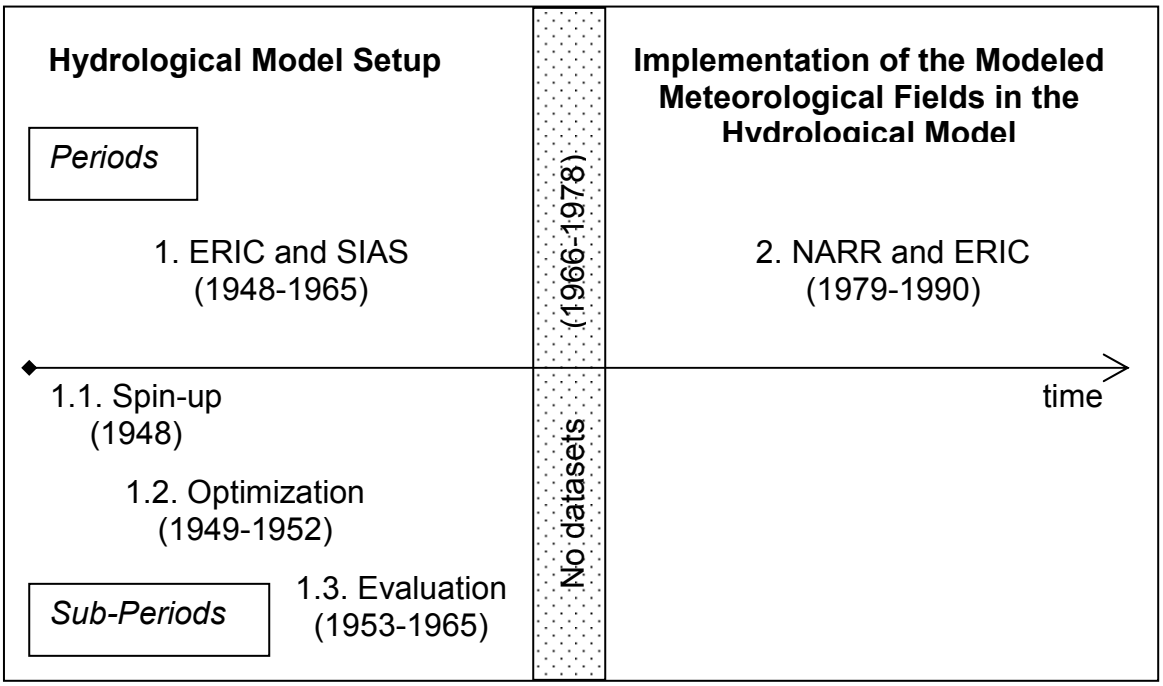


Figure 3.1. Modeling timelines used in this study. The panel above the arrow indicates periods when data were available (in parenthesis), the use made of these data (bold letters), and the datasets calculated using data from ERIC applied as forcing to the model available. The panel below the arrow indicates sub-periods and data use.

4. RESULTS

4.1. Setting Up MMS-PRMS

A rectangular area was defined which contained the *Río Grijalva* basin, with the upper left corner at 94.5°W; 19°N and lower right corner at 91°W; 14.5°N (Fig 4.1), and data from all the 356 meteorological stations inside this rectangle extracted from ERIC and combined with that from INSIVUMEH stations. Data from these 356 stations is intermittent: the greatest number of stations reporting on one day was 267 and the lowest number 93 in the period 1948-1965.

Data from hydrologic stations for the basin were also extracted from SIAS (Fig 4.2). Selection of the target (outlet) station for modeling was based on the quality of the SRTM grid and the location of the streamflow stations. The floodplain which is downstream of station 30015 contains several water bodies that are not well defined by SRTM, and this lead to the specification of an unrealistic river-network by GIS-Weasel. This floodplain was therefore excluded from the present study. Modeling this region will require a digital topography grid of higher quality and better resolution. Station 30015, on the other hand, is the only gauging station that aggregates all the sub-basins upstream of the floodplain and for this reason was selected as the target station. The corresponding watershed contour derived by GIS-Weasel from digital topography is shown in Fig 4.3.

The topography grid was also used with GIS-Weasel to define the configuration of the HRUs. Different area thresholds give different HRUs-

configurations, and the HRU configuration obtained with a threshold of 259.2km^2 comprising 32,000 cells each 90m by 90m was selected. With this threshold, the nodes of the HRUs for the most part concur with the location of the dams (Fig. 4.4) and this means their individual modeling would be possible, if required. The river network was also derived from the digital topography and the topographic, soil, and vegetation parameters required by the model then derived for each of the selected HRUs.

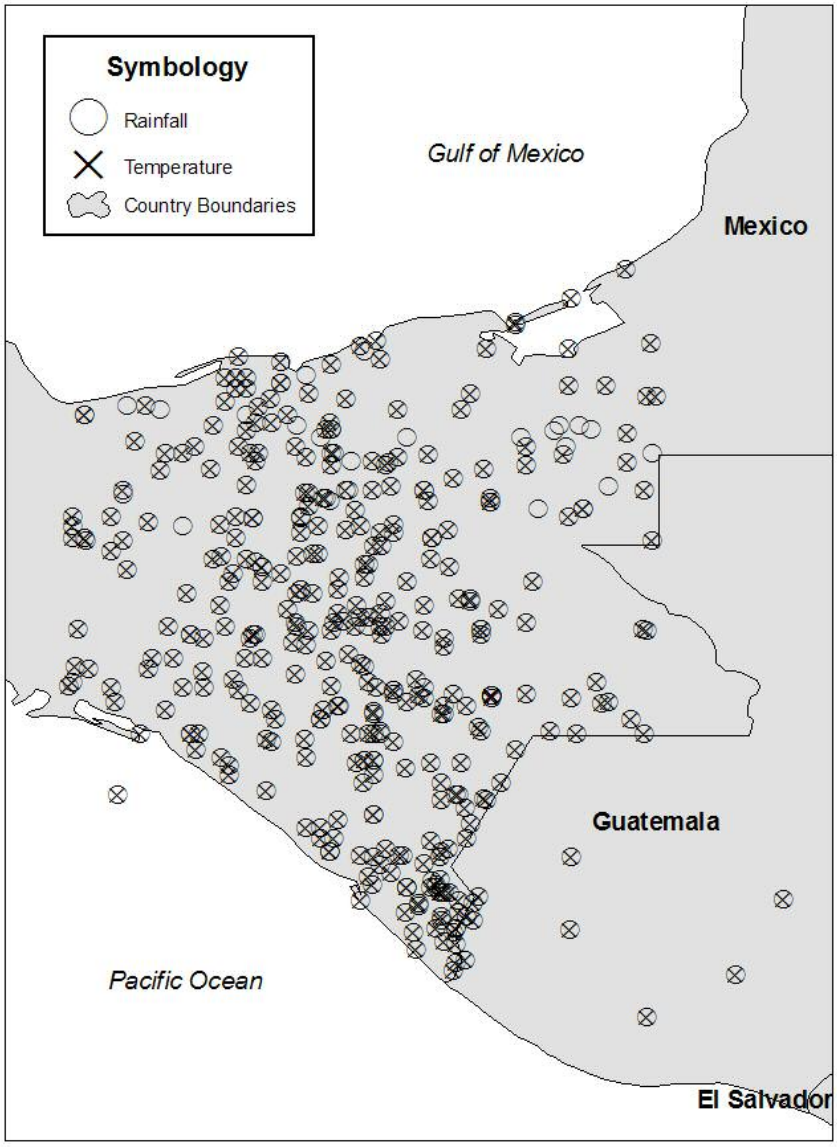


Figure 4.1. Map of meteorological stations. Circles indicate stations reporting rainfall, while crosses indicate stations reporting temperature.

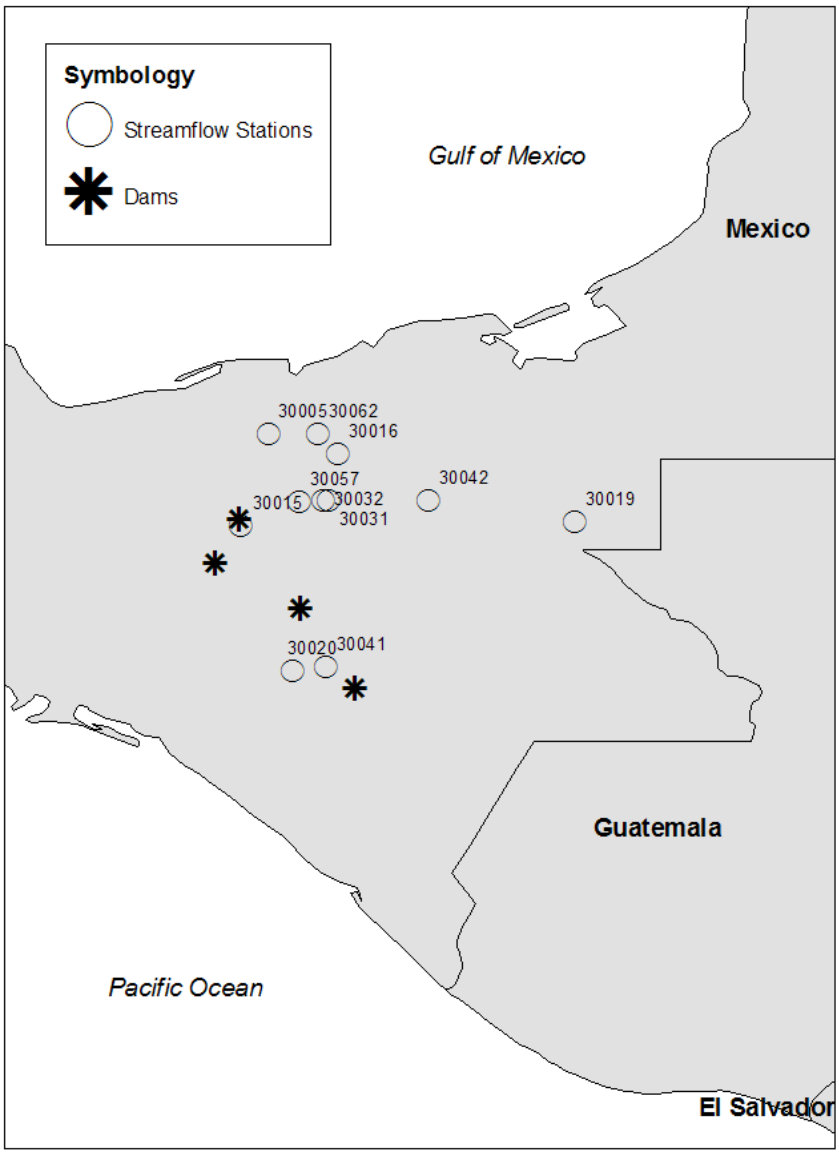


Figure 4.2. Map of hydrological stations (circles) and dams (stars). Numbers associated with circles represent the station index in SIAS.

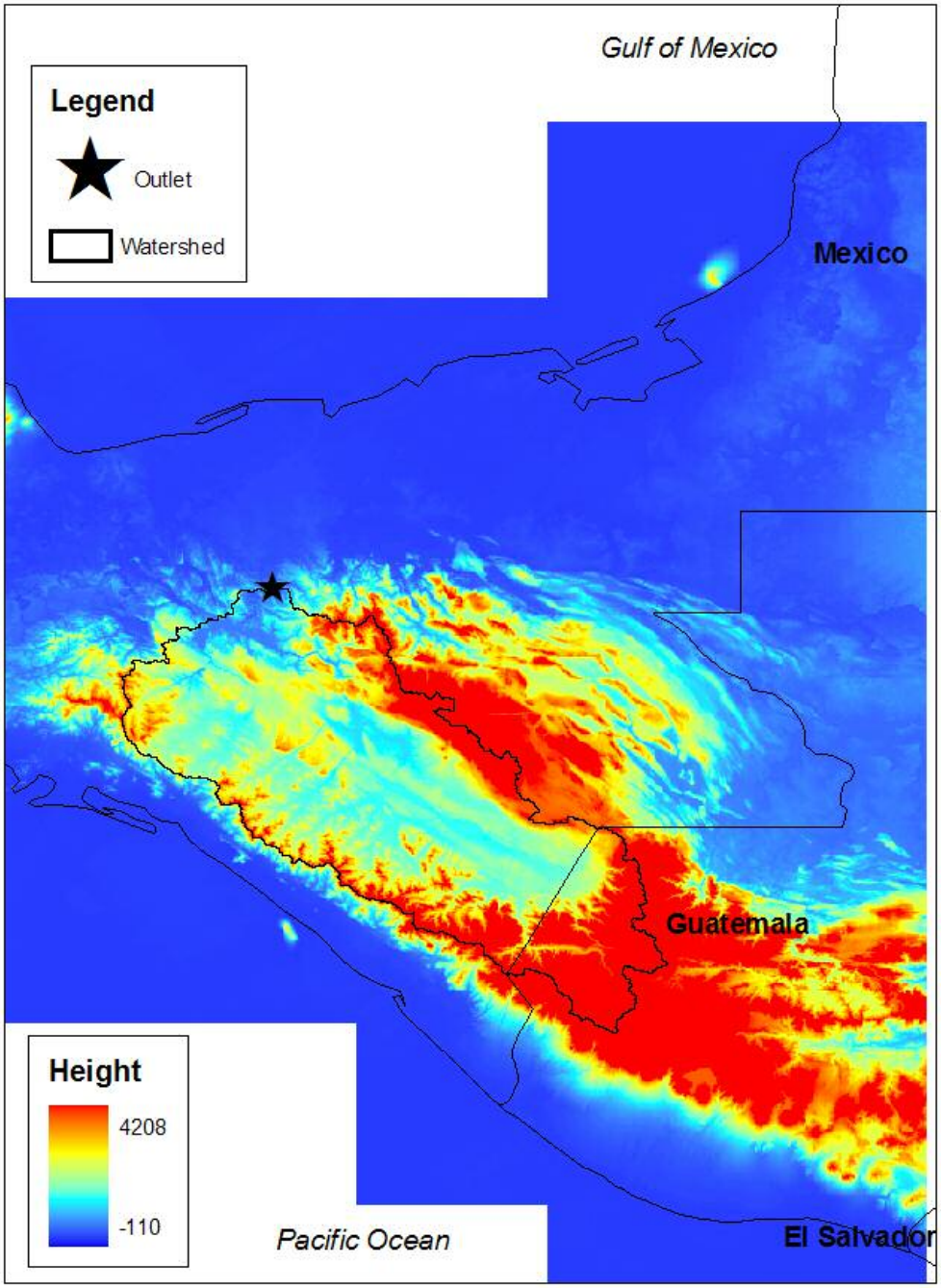


Figure 4.3. Height above mean sea level [m], and watershed contour (thick black line) for the target outlet (star) derived by GIS-Weasel.

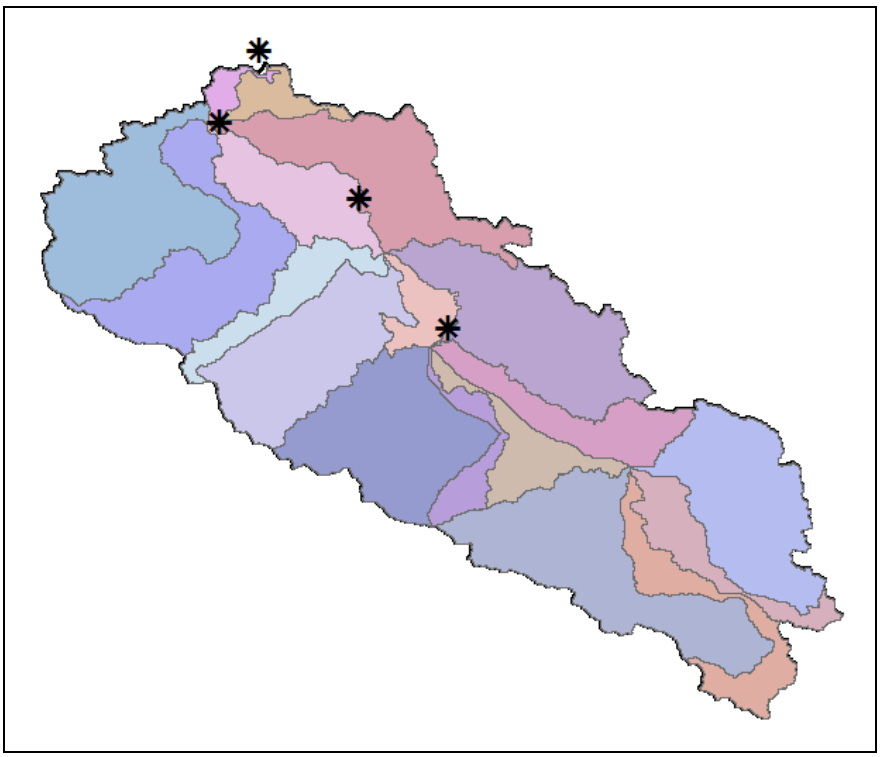
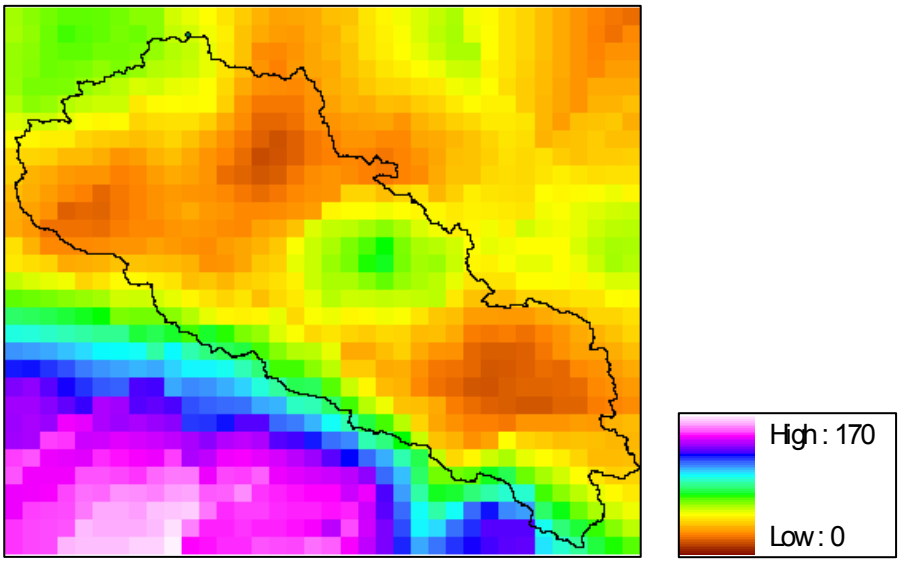


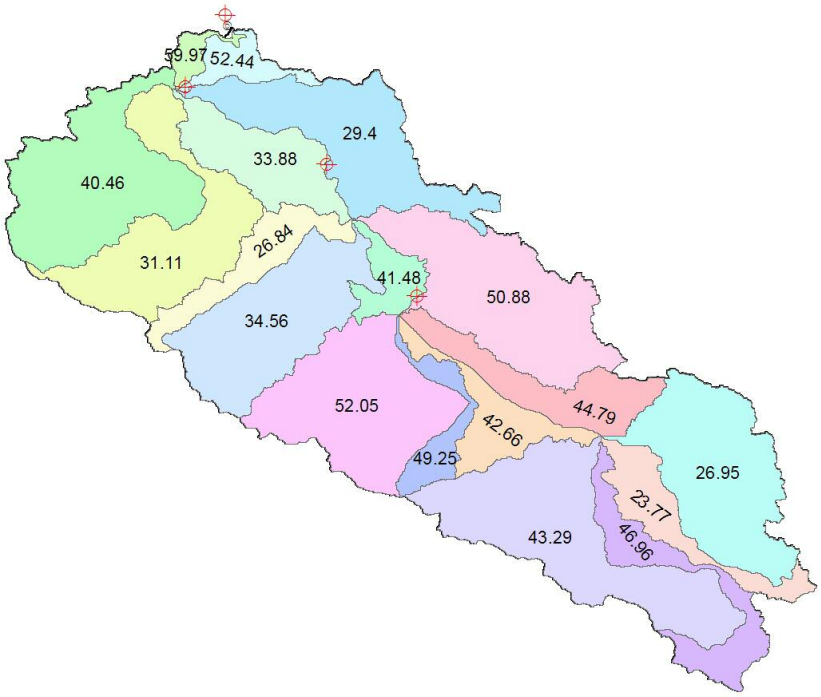
Figure 4.4. Map of Hydrological Response Units (HRUs). The stars indicate the location of the dams in the basin.

Daily meteorological data from the available stations are interpolated using a kriging technique. The meteorological fields so obtained were found to realistically reflect the presence of important features such as the topography and ocean-continent boundaries. In this region, the rainy-season is associated with a significant flow of humidity from the Pacific Ocean. When this flow meets the mountain range (the *Sierra Madre de Chiapas*) near the southwest edge of the watershed, orographic lifting generates strong rainfall. This is why the interpolated rainfall grids show a strong gradient at the southwest boundary of

the basin, see Fig. 4.5(a). In order to interface the interpolated meteorological fields with the hydrological model, the fields were then projected over each HRU and the relevant cells inside every HRU arithmetically averaged. Figure 4.5(b) shows sample average values for all the HRUs.

(a)





(b)

Figure 4.5. (a) Sample interpolated field of observed rainfall (September 19th, 1978) derived using kriging [mm]; and (b) the average value of rainfall derived from this interpolated field for each HRU [mm].

The same process was used to generate an input file of daily maximum and minimum temperatures and rainfall from 1948 to 1965. Streamflow modeled by MMS-PRMS using this dataset is shown in Fig. 4.6. The modeled streamflow captures the general trend of the observed streamflow for the whole period, but a noteworthy deficiency is the consistent underestimation of the recession during winter and spring (more evident in the plot with logarithmic scale). In fact, the plot for the selected period 1949-1952 which is shown in Fig. 4.7 shows that this deficiency during the winter recession is preceded by an overestimation of the runoff. Additionally, the plot for year 1951, which is shown Fig. 4.8, suggests that much of the intra-seasonal variability is missing in the modeled time series.

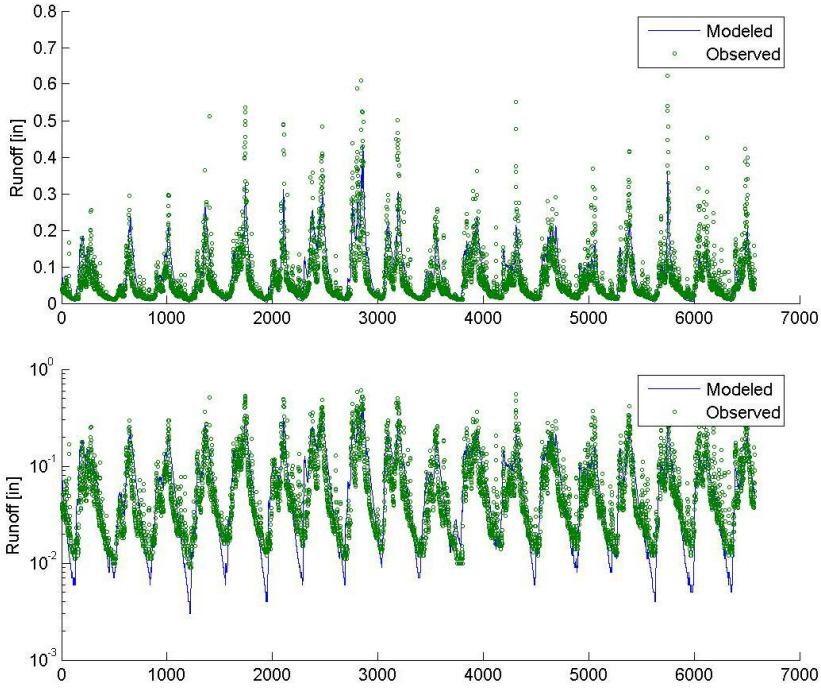


Figure 4.6. Daily observed streamflow compared with that modeled using data from meteorological stations [in] calculated using the model prior to optimization for period 1948-1965 plotted using linear (upper panel) and logarithmic (lower panel) scales.

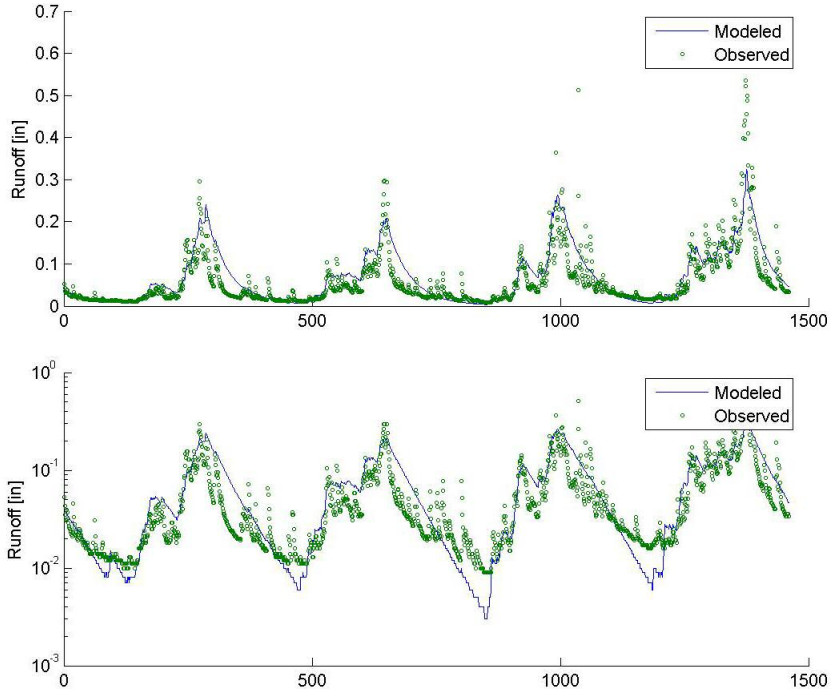


Figure 4.7. As for Fig. 4.6 but for the period 1949-1952.

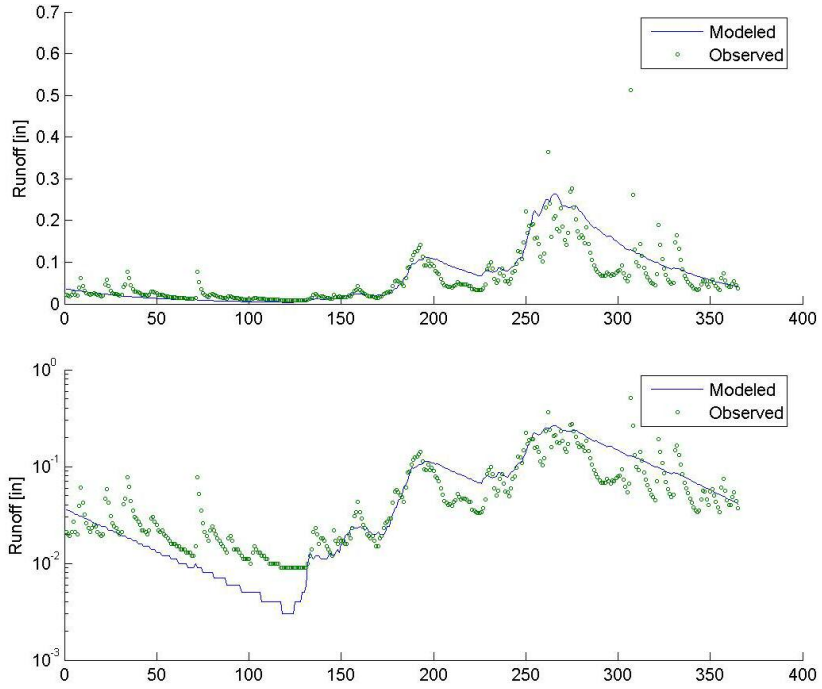


Figure 4.8. As for Fig. 4.6 but for 1951.

The initial model performance was improved by optimization with LUCA, using the procedure described in section 3.4. By the end of the optimization process, the value of the Objective Function (OF) was reduced to 41% of the initial value, see Table 4.1. The greatest reduction in OF was occurred after the first two optimization steps. Subsequent efforts to optimize the modules relating to interception and evaporation resulted in a negligible improvement in model performance, consequently prescribed values of relevant parameters were used in these two modules.

Table 4.1. Evolution of the Objective Function value during optimization.

Round	Step	Modules	Number of Model Evaluations	OF value	Percentage of the initial value
0	*0*	-	*0*	*696.00*	*100.00*
1	1	*gwflow and smbal*	302	340.00	48.85
1	2	*ssflow and smbal*	726	301.83	43.37
1	3	*srunoff and smbal*	376	301.58	43.33
1	4	*intcp and smbal*	-	-	-
1	5	*potet and smbal*	-	-	-
2	1	*gwflow and smbal*	874	297.00	42.67
2	2	*ssflow and smbal*	886	288.00	41.38
2	3	*srunoff and smbal*	-	-	-
2	4	*intcp and smbal*	-	-	-
2	5	*potet and smbal*	-	-	-

Comparison of the modeled and observed streamflow indicates that baseflow is one of the aspects of the simulation that most benefited from optimization, see Figures 4.9 and 4.10. Overestimation in the earlier portion of

the dry-season and underestimation after the dry season is no longer observed. It is likely that the sequential order of the modules optimized and the selection of PNAD, the Objective Function, is responsible for the improvement in this low-frequency component of the hydrograph. However, some the intra-seasonal variability is still missing in the modeled streamflow, particularly the peak flows during the winter recession.

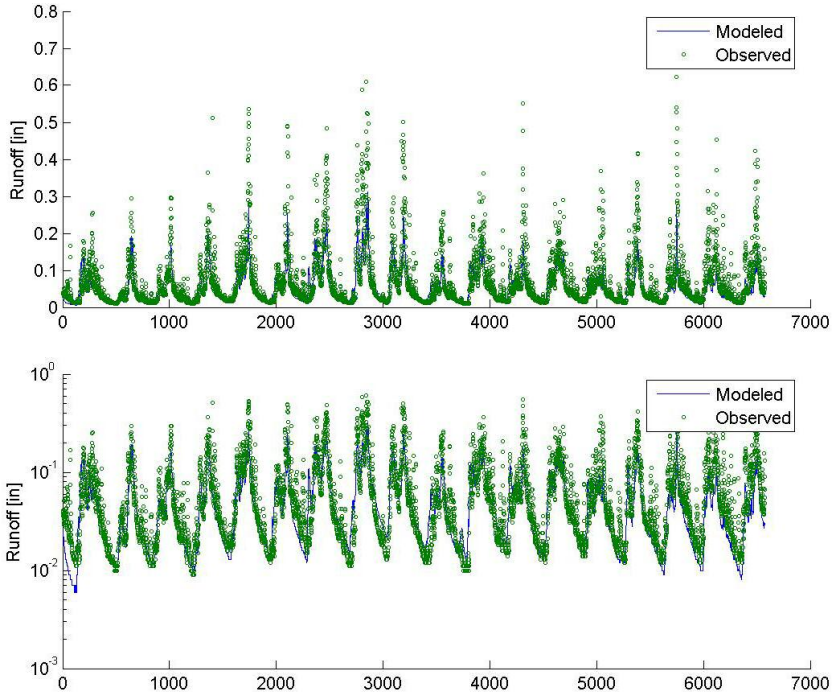


Figure 4.9. Daily observed streamflow compared with that modeled using data from meteorological stations [in] calculated using the model after optimization for period 1948-1965 plotted using linear (upper panel) and logarithmic (lower panel) scales.

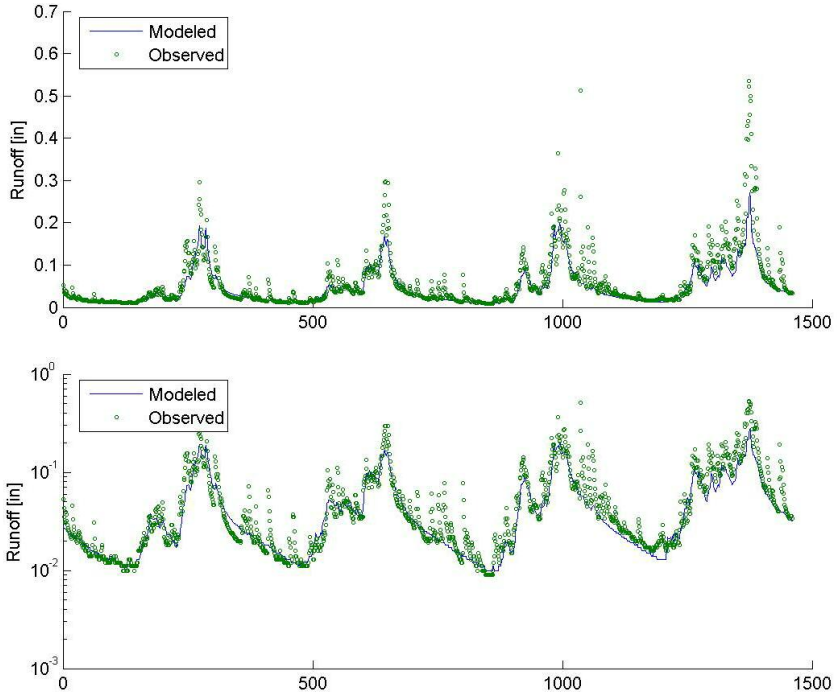


Figure 4.10. As for Fig. 4.9 but for the period 1949-1952.

Table 4.2 documents that their was a substantial reduction in an objective function similar to the one used, i.e. the linear mean absolute difference (MAD), in all data periods following optimization. The logarithm of this objective function (the mean logarithmic absolute difference (MLAD), also shows a significant reduction but a quadratic objective function (the MSE) was little improved during the optimization period and in fact was worsened during the evaluation period.

Table 4.2. Objective Functions (Mean Logarithmic Absolute Difference (MLAD), Mean Absolute Difference (MAD), and Mean Square Error (MSE)) and their decrease after optimization for the different periods involved in setting up the hydrological model.

Period	Model	MLAD	Decrease (%)	MAD	Decrease (%)	MSE	Decrease (%)
Spin-up (1948)	Non-Optimized	0.0343		0.7806		0.0026	
	Optimized	0.0220	35.9	0.4408	43.5	0.0012	53.8
Optimization (1949-1952)	Non-Optimized	0.0233		0.4770		0.0014	
	Optimized	0.0157	32.6	0.2066	56.7	0.0013	7.1
Evaluation (1953-1965)	Non-Optimized	0.0290		0.5212		0.0020	
	Optimized	0.0245	15.5	0.3348	35.8	0.0023	-15.0

Additional modeled components of the hydrological cycle are shown in Fig. 4.11 for 1955 and and Fig. 4.12 for 1960, these being the two years for which the MLAD was highest and lowest for the optimized model, respectively. The figure confirms there are small changes in the evapotranspiration, perhaps associated with the fact that the parameters associated with this process were not optimized. It is interesting that although soil moisture was calibrated at every step during optimiziation, the total storage below the surface (soil, subsurface and groundwater storages) were little changed by the optimization process. Optimization makes most modification to the modeled streamflow, this being strongly related to groundwater and subsurface flow.

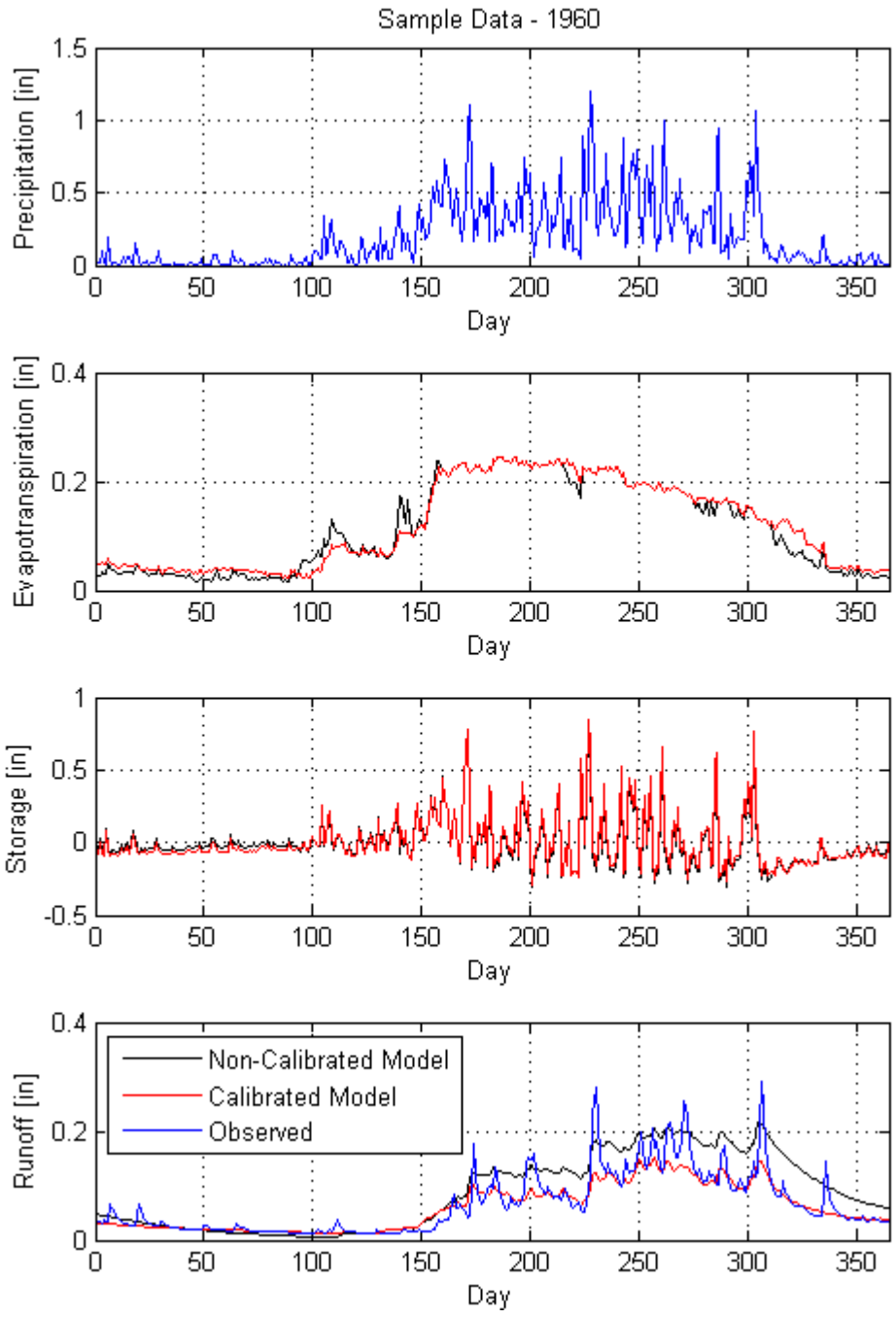


Figure 4.11. Precipitation, evapotranspiration, total storage, and runoff for 1960. Observations are indicated in blue, output from the calibrated model are shown in red, and output from the non-calibrated model is shown in black.

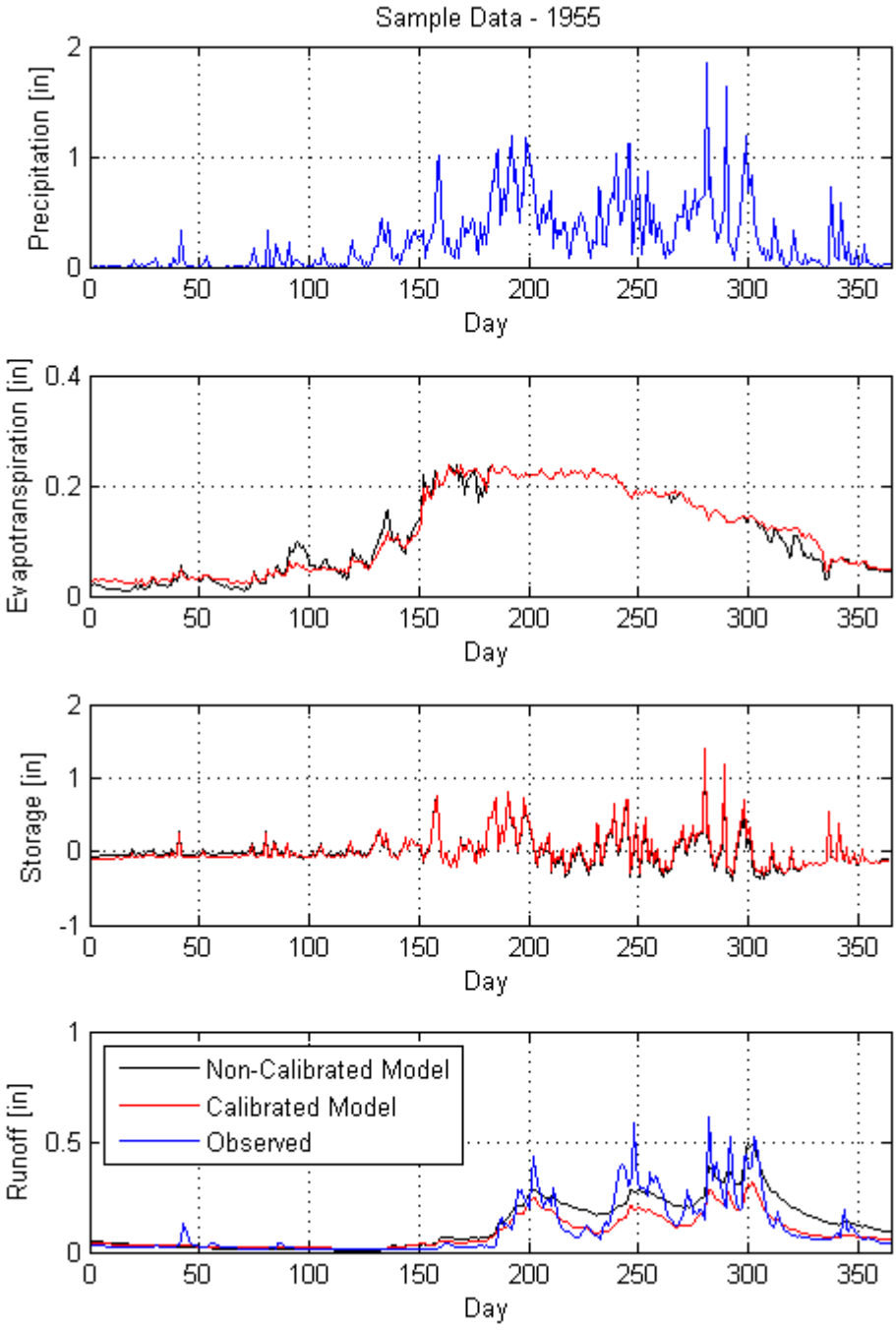


Figure 4.12. As for Fig. 4.11 but for 1955.

Flow Duration Curves (FDCs) for the observed, un-optimized, and optimized models for both the rising and falling limbs of the streamflow shown in Fig. 4.13 demonstrate that there is closer resemblance between observed and modeled streamflow after optimization. Simulation of the rising limb is least benefited by model optimization.

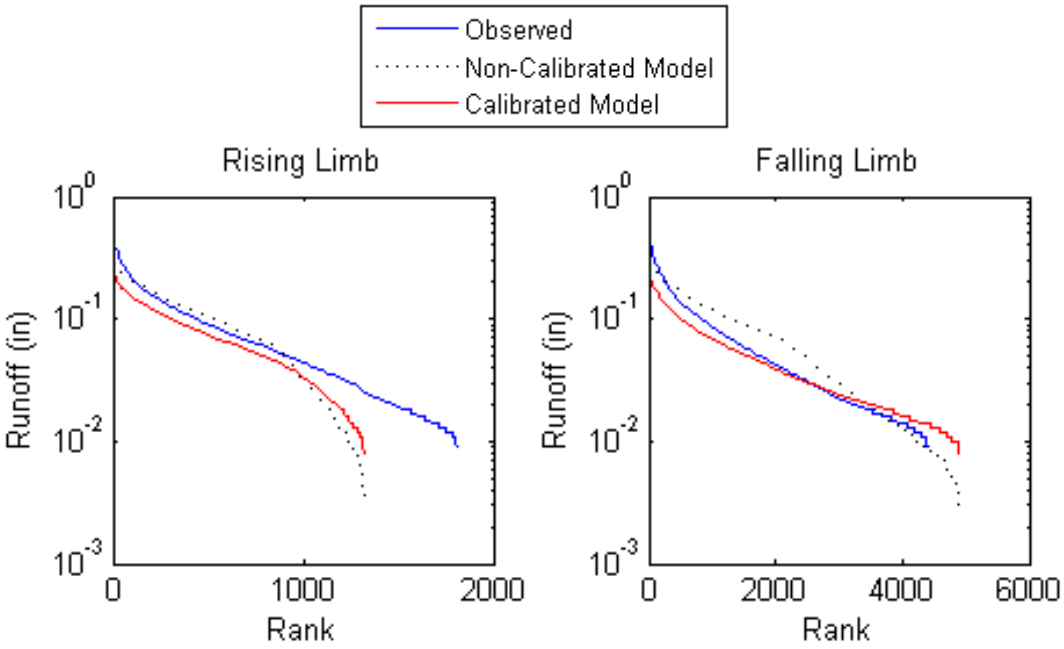


Figure 4.13. Flow Duration Curves for the observed (blue) and modeled streamflows calculated using unoptimized (dotted black line) and optimized models (red line) for both the rising limb (left plot) and falling limb (right plot).

4.2. Modeling Using NARR Data

NARR data fields of temperatures and rainfall were used to force the optimized model and the resulting modeled streamflow is compared to the modeled streamflow from raingage data (Fig. 4.14). The results indicate that the model greatly underestimated streamflow when forced with NARR data and since it is the rainfall forcing which is most likely the cause of this underestimation, a detailed investigation of rainfall bias in the NARR data was therefore undertaken.

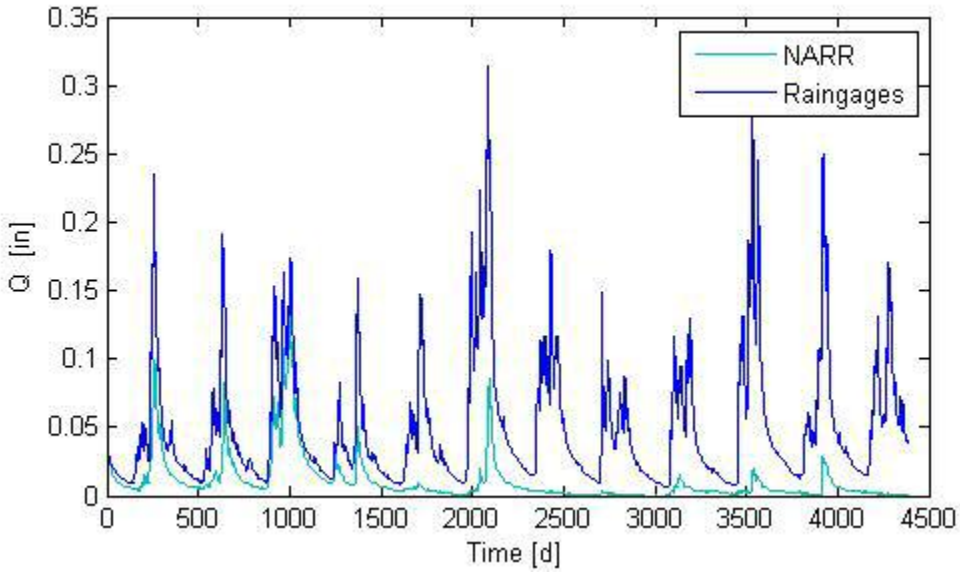


Figure 4.14. Daily modeled streamflow calculated with the optimized hydrological model using data derived from the meteorological stations (dark blue) and the NARR fields (light blue) for period 1979-1990.

The average rainfall depth for the whole basin was used in analysis described below. Cumulative NARR rainfall is only 65% of the observed rainfall (Fig. 4.15) but there is no evidence of a simple systematic bias when the daily

rainfall amounts are compared (Fig. 4.16). However, a systematic lag is apparent in the correlation coefficient between observed rainfall and NARR rainfall calculated for different lag-times (Fig. 4.17). The maximum correlation coefficient (0.71) occurs when NARR-rainfall is delayed by one day, and this lag was therefore applied to this dataset prior to its use in the hydrological model. The analysis of the bias distribution shown in Fig 4.18 is interesting: the NARR bias is within +/-1mm around 47% of the time, but there nonetheless is a strong tendency towards rainfall underestimation. The Root Mean Square Error (RMSE) is 5.11mm, and no-rain days are well forecast only 31% of the time while rainy days are well forecasted 95% of the time. Clearly there are obvious deficiencies in NARR rainfall data, but some aspects suggest the possibility of improvement using an appropriate bias correction methodology.

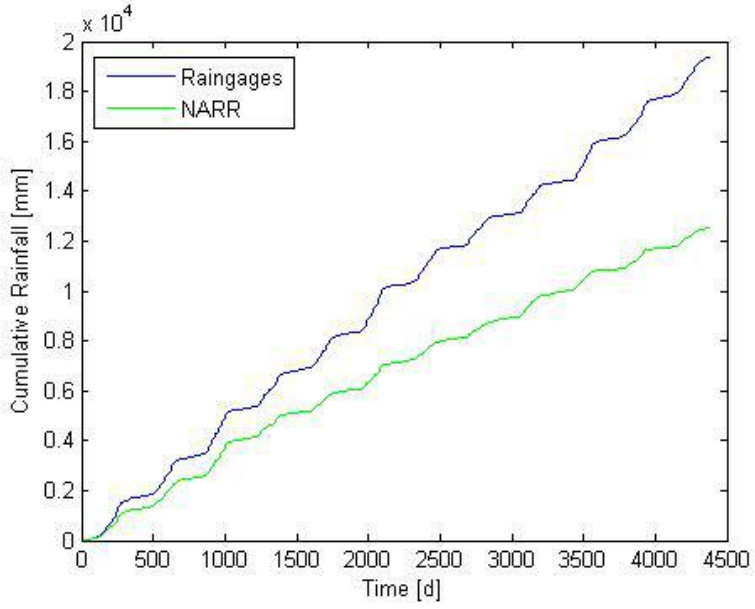


Figure 4.15. Cumulative rainfall-depth over the whole basin for both rain-gauges (blue) and NARR rainfall (green).

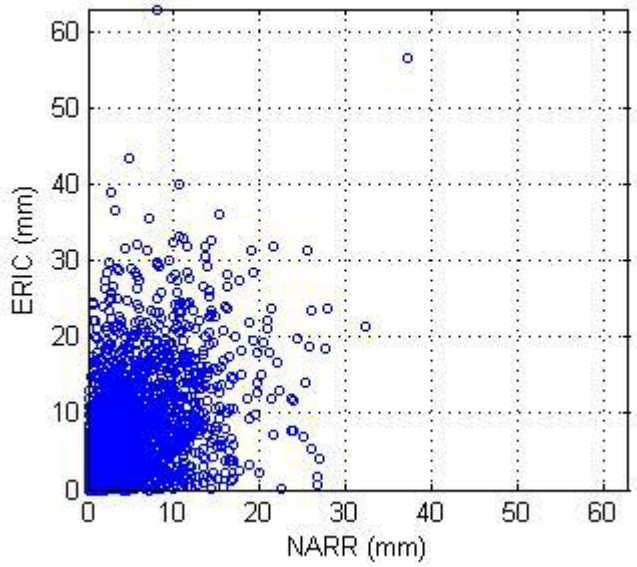


Figure 4.16. NARR rainfall versus observed rainfall (from ERIC).

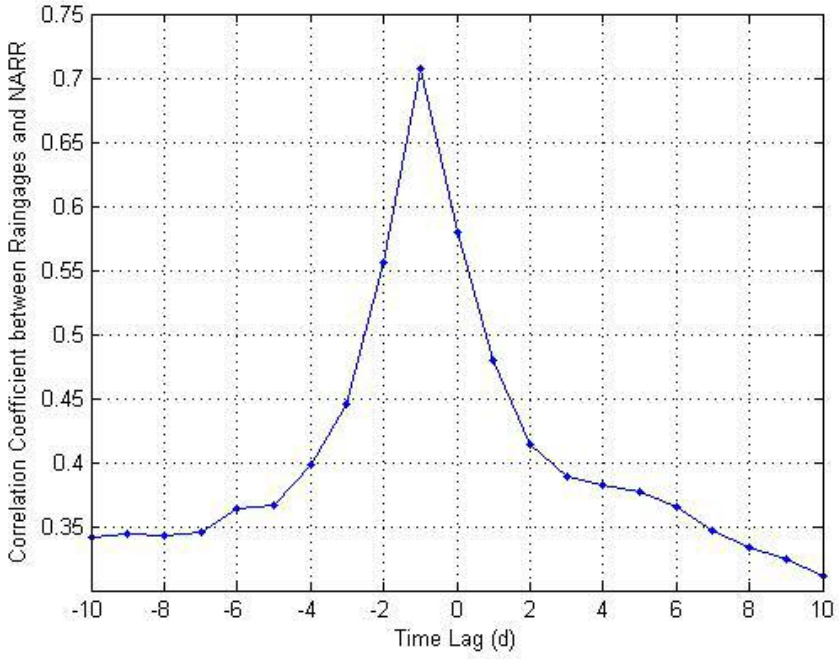


Figure 4.17. Correlation coefficient between raingages and NARR-rainfall. Negative (positive) lags indicate a delay (advance) in the NARR-rainfall.

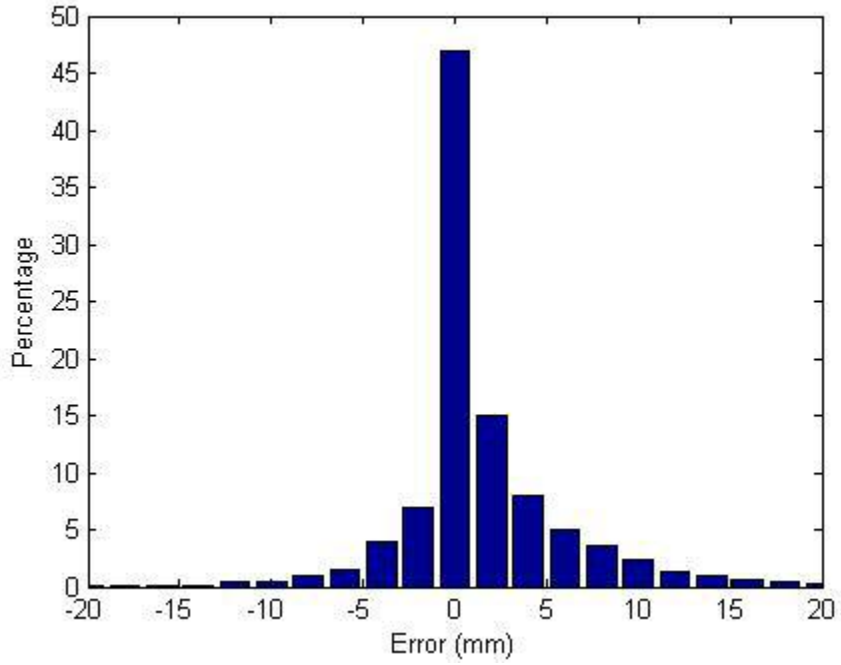


Figure 4.18. Histogram of bias in the daily rainfall calculated as observed rainfall minus NARR-rainfall amount.

The most important factor in the definition of a correction methodology is the nature of the observed biases. The plot of bias as a function of time indicates heteroscedasticity (Fig. 4.19), which is defined as the presence of different standard deviations in a single random variable. The standard deviations of the biases are larger during the rainy season (strong events) and smaller in the dry-season (weak events). Because heteroscedasticity makes the application of linear correction methodologies impractical, alternative correction methods were explored.

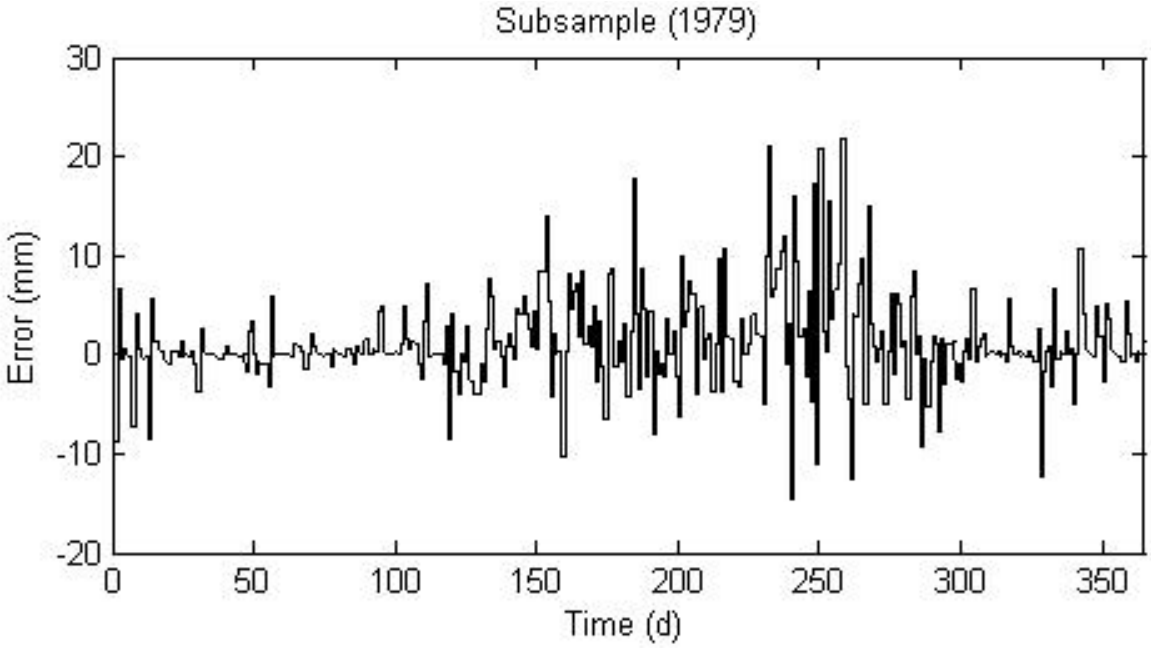


Figure 4.19. Sample plot of daily bias (observed rainfall minus NARR-rainfall) versus time for 1979. Start and end-dates of the rainy season are usually around May (days 120-150), and September (days 240-270), respectively.

4.2.1. Bias Correction Methodology based on the Modeled-Observed Joint - Probability Distribution Function (MOJ-PDF)

The Modeled-Observed Joint - Probability Distribution Function (MOJ-PDF) for the seasonal rainfall is the critical component in the definition of the bias correction method proposed applied in the present study. The procedure used in its derivation involves first assigning the NARR rainfall estimates into regular percentile bins. Twenty 5% bins were used in this study, see the upper panel of

Fig. 4.20. The PDFs for rainfall observations that corresponded to the NARR rainfall estimates assigned to each bins were then derived. The lower panel in Figure 4.20 shows the median (50-percentile) and extreme percentiles (5 percentile and 95 percentiles) for each bin. This analysis yields the required sampling of the MOJ-PDF.

The first conclusion that can be drawn from the MOJ-PDF analysis described above is that during both the dry and rainy seasons, the smallest but most frequent rainfall events tend to be underestimated by the NARR rainfall data while the extreme events tend to be overestimated. The net underestimation is larger in the rainy season and this deficiency is therefore largely responsible for the difference between cumulative NARR rainfall and the observed rainfall. The MOJ-PDF analysis also confirms the presence of and quantifies the heteroscedasticity, because stronger (weaker) events have larger (smaller) statistical dispersions. Consequently, a correction methodology based on the dispersion and the season of the event is to be preferred over a linear, deterministic approach (e.g., a correction factor or a linear-regression), and a probabilistic approach based on the MOJ-PDF was therefore explored.

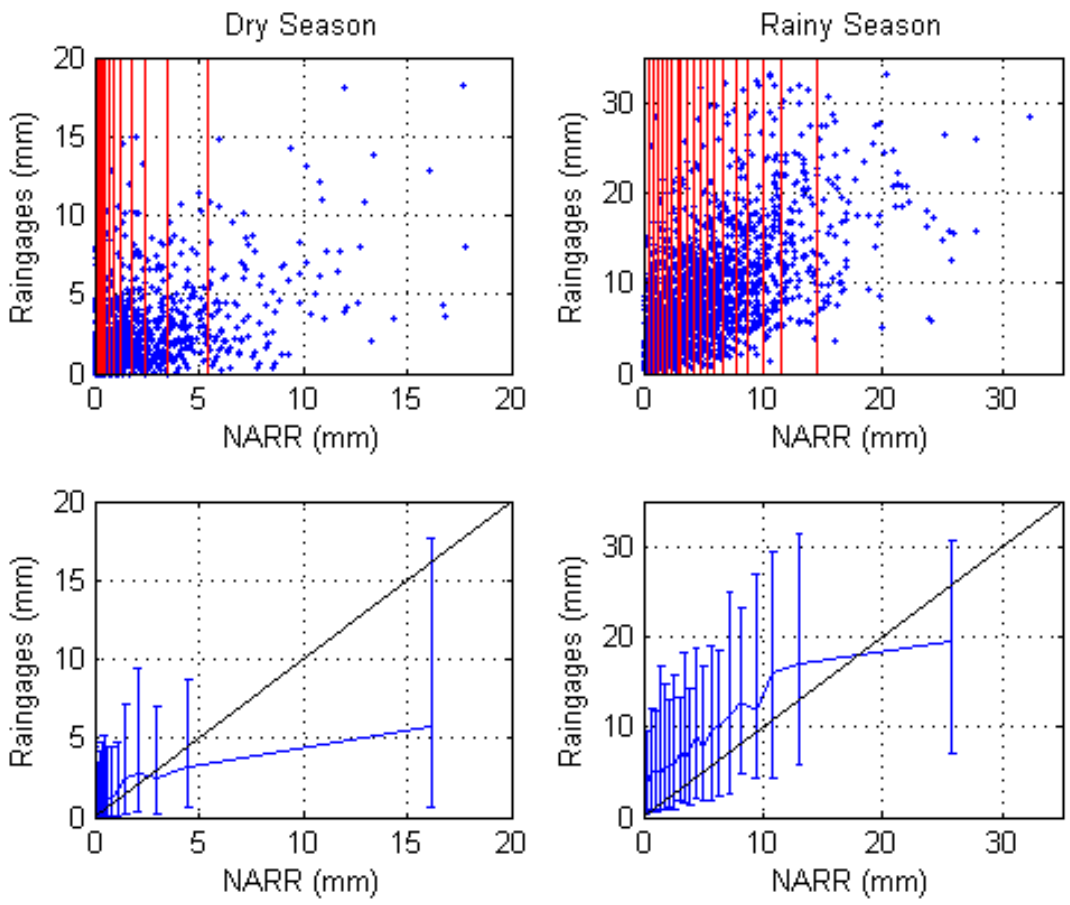


Figure 4.20. Observed versus modeled rainfall (blue dots), and 5-percentile bins (limited by red lines) based on NARR-rainfall are shown in the upper panels. The lower panels show the medians (50-percentile) significant extreme percentiles (5 and 95) for each of these bins and 1:1 (black) line. Plots on the left correspond to the dry season, plots on the right correspond to the rainy season.

One obvious correction procedure based on MOJ-PDF analysis would be to replace a given modeled event by the value of the median (50-percentile) for the bin in which the NARR rainfall estimate falls. In practice applying this method did decrease the difference between cumulative modeled and observed rainfalls from 35% to 17% (Fig. 4.21), but the difference is still large. Therefore an

alternative approach was used in which the original values were replaced by the value corresponding to the percentile which, on average, nullifies the difference between observed and NARR rainfall, for this dataset. This is the 61 percentile. The corrected average rainfall over the whole basin is used to estimate the corrected values for every HRU assuming the correction factor for the whole basin applies to each HRU.

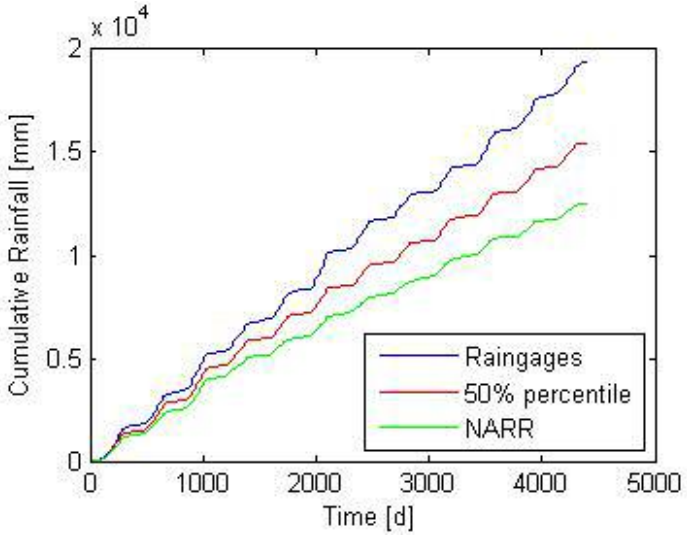


Figure 4.21. Cumulative rainfall-depth over the whole basin for rain-gauges (blue), the median of the MOJ-PDF (red), and the original NARR (green).

The NARR rainfall time series was corrected in this way and the result used as forcing for the hydrological model. This resulted in a noteworthy improvement in the level of agreement between the streamflow modeled with NARR data and observed data: compare Fig. 4.22 with Fig. 4.14. The Flow Duration Curves show a similar improvement, see Fig. 4.23.

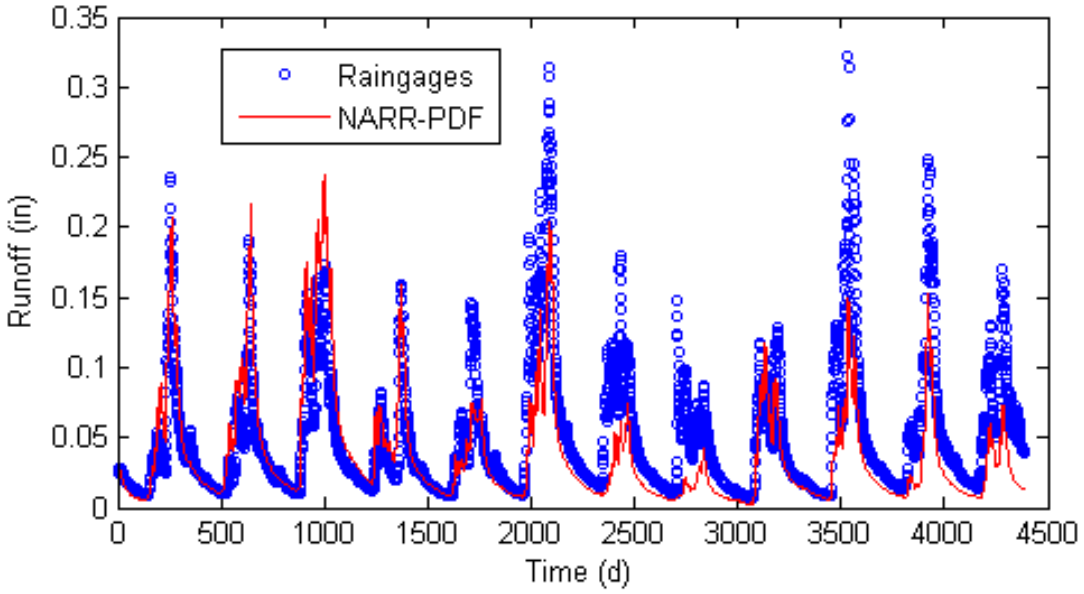


Figure 4.22. Daily modeled streamflow calculated using raingage data (blue dots) and using NARR rainfall data corrected using the MOJ-PDF based methodology (red line).

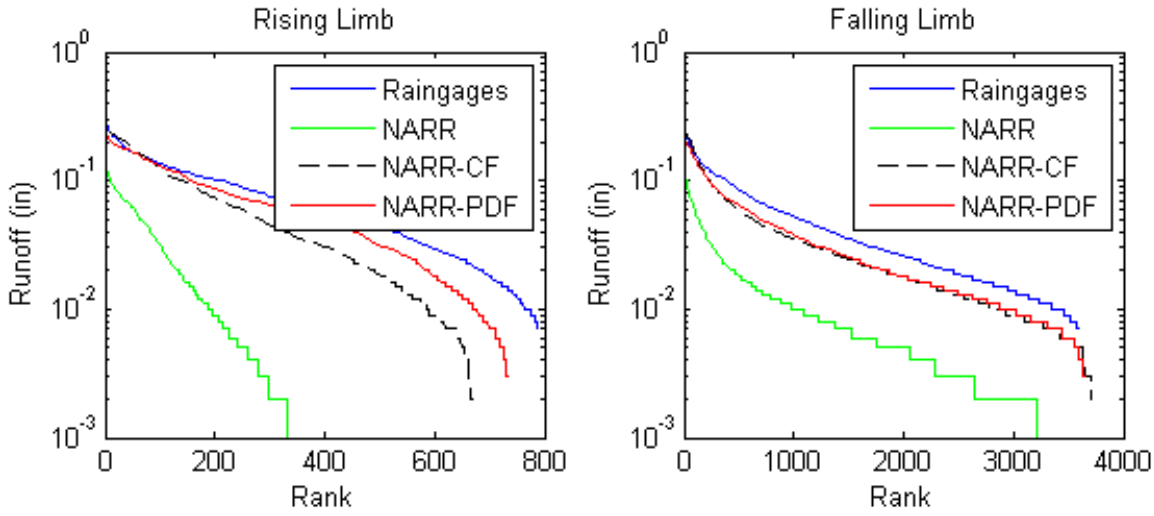


Figure 4.23. Flow Duration Curves for the modeled streamflow calculated from raingages data (blue), the original NARR rainfall data (green), NARR rainfall corrected by a single correction factor (black), and NARR rainfall data after correction using the MOJ-PDF based methodology (red).

4.2.2. Bias Correction Methodology based on the Maximum Likelihood Estimator (MLE)

As a validation of the MOJ-PDF based correction methodology, a second correction approach was also applied to the NARR rainfall data based on a Maximum Likelihood Estimator (MLE (Sorooshian and Dracup, 1980)). The method involves applying a transformation that stabilizes the variance, and in this study the following power transformation was selected:

$$y = \frac{I^{\lambda} + 1}{\lambda},$$

where *I* is the rainfall and λ a parameter whose value determines the shape of the power transformation. The variation of *y* as a function of *x* for different values of λ is shown in Fig. 4.24.

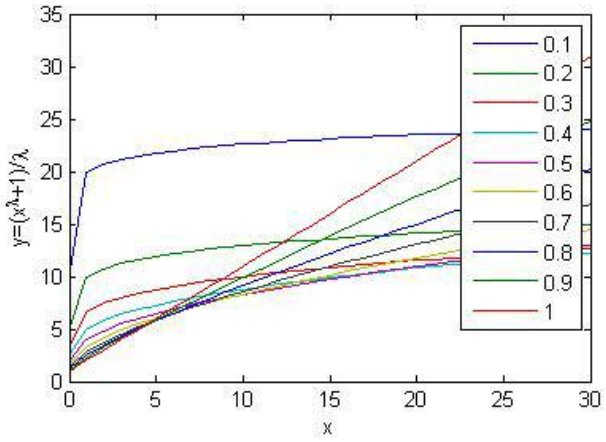


Figure 4.24. Power transformation of the Maximum Likelihood Estimator for different values of the parameter λ.

The procedure used to make this MLE correction is illustrated in Fig 4.25. The first step in the procedure is to identify the value of λ that gives the highest correlation coefficient between the variable to be corrected, in this case the NARR-rainfall, the independent variable, and the observed rainfall, the dependent variable. The transformation thus defined is then applied to the variables and linear regression used to identify the linear-relationship between transformed variables. The resulting relationship is then used to generate a third rainfall time-series which corresponds to the “transformed observed rainfall” that gives the minimum error. The time series is then inversely transformed to return a rainfall estimate. Finally, a correction factor is applied that closes the gap between cumulative NARR rainfall and the observed cumulative rainfall. Again the average correction factor for the whole basin is used to calculate the corrected values for each HRU.

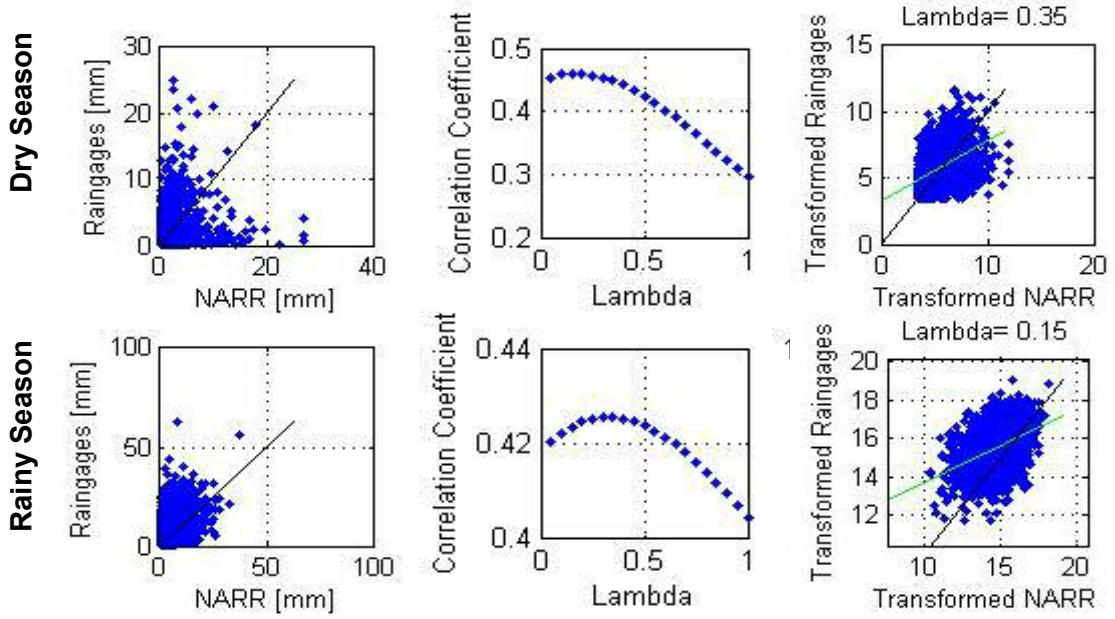


Figure 4.25. Illustration of procedure used to correct NARR rainfall using the Maximum Likelihood Estimator methodology. The upper panels relate to the dry season and lower panels to the rainy season. From left to right the diagrams show (1) the scattergram for non-transformed variables; (2) the correlation coefficients for different values of λ, and (3) the scattergram for transformed variables with the linear regression line (in green) and the 1:1 line (in black).

In practice, both MOJ-PDF and MLE correction approaches give very similar results in terms of modeled streamflow (Fig. 4.26). However, the MOJ-PDF analysis can be easily extended to address uncertainty issues in the NARR rainfall, and it was preferred in this study for this reason.

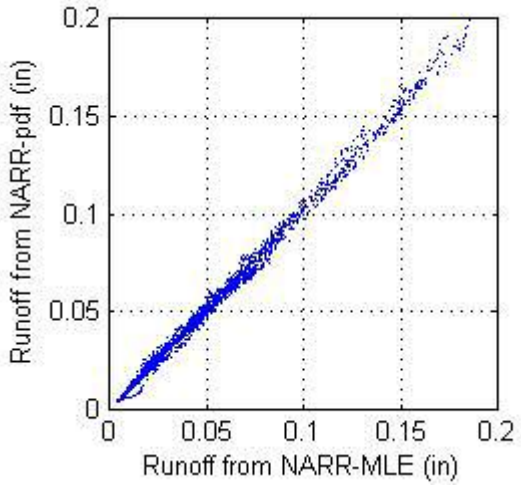


Figure 4.26. Modeled streamflow calculated with NARR-rainfall after correction with the MOJ-PDF method relative to that when calculated with NARR rainfall corrected using the MLE method.

4.3. Uncertainty Analysis of NARR Rainfall

The method used to address uncertainty in the NARR rainfall estimate is largely based on the MOJ-PDF bias correction methodology previously described. However, instead of selecting a single value (i.e. the percentile 61) to correct the NARR rainfall, several values of percentile are randomly sampled from the MOJ-PDF within the confidence intervals commonly employed in uncertainty analysis, namely 5% to 95%. (Note: it is recognized that this method approach cannot fully preserve the temporal correlation between observed and modeled rainfall).

Figure 4.27 illustrates the procedure used to estimate the uncertainty in the corrected NARR rainfall for August 30th, 1983 (this day was selected at random). The NARR rainfall estimate lies in the bin limited by percentiles 85 and 90 on this day; see the upper panel of this figure. The middle panel shows the

probability distribution of observed rainfall that corresponds to this NARR rainfall bin. Percentage values are randomly selected a number of times, one-hundred times in this case, assuming a uniform distribution across the range between the values 5% and 95%. The observed rainfall corresponding to each sampled percentage is defined from the distribution of observed rainfall for this particular NARR rainfall bin. In this way a 100 member sample of observed rainfall estimates was created which is drawn appropriately from the observed values relevant to the NARR rainfall estimate. The lower panel of the figure shows the distribution of such randomly generated rainfall events. This resembles the original distribution of observed rainfall within the selected confidence intervals, as it should, and confirms the methodology had been implemented correctly.

The procedure described above was applied to the daily NARR-rainfall dataset from 1979 to 1982. Figure 4.28 shows the resulting one hundred time series for a sample year (1979). The hydrological model was then forced with each of these one hundred input datasets to generate a 100-member ensemble of runoff time series. On each day, the upper-limit (lower-limit) of the uncertainty region was defined as the maximum (minimum) streamflow. Fig. 4.29 shows that the modeled streamflow calculated using raingage data lies within this uncertainty range most of the time suggesting the methodology is estimating the uncertainty of the NARR rainfall appropriately.

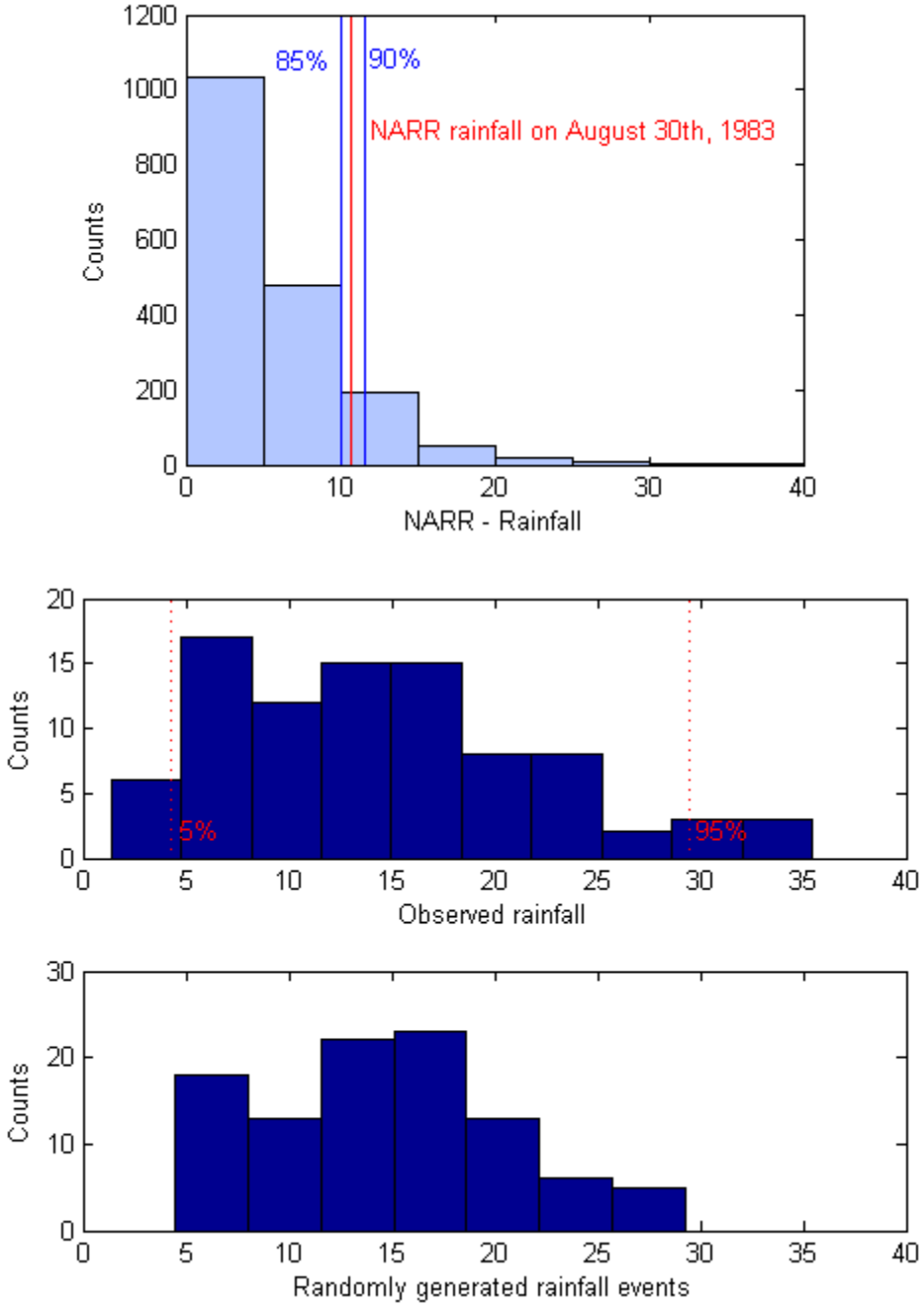


Figure 4.27. Histogram of NARR-rainfall during the rainy seasons (upper panel). The red line indicates modeled rainfall for a given event (August 30th, 1983). Blue lines indicate the limits of the corresponding 5-percentile bin. Histogram of observed rainfall for the selected bin (middle panel). Red lines indicate 5 and 95 percentiles. Histogram of randomly generated rainfall (lower panel).

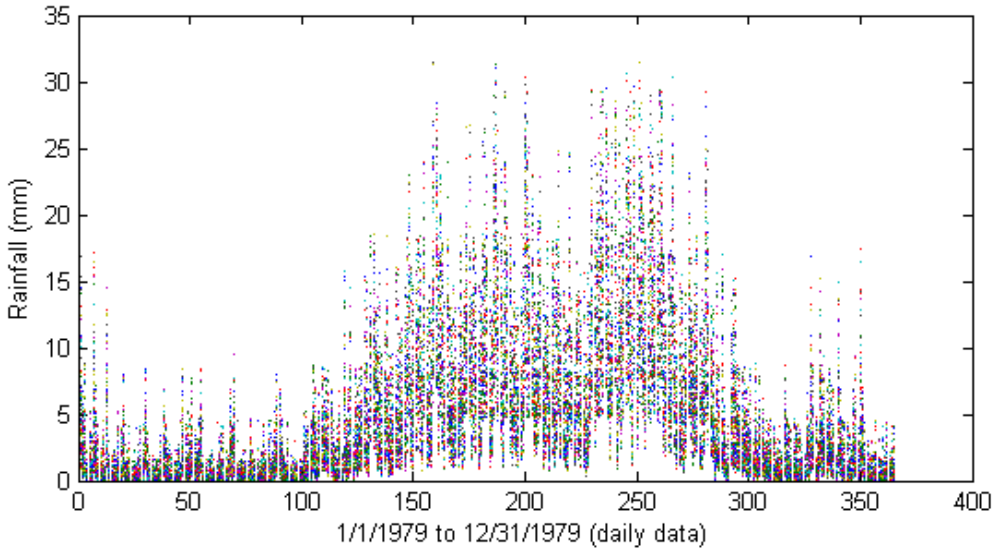


Figure 4.28. One-hundred randomly generated time series of daily rainfall for year 1979.

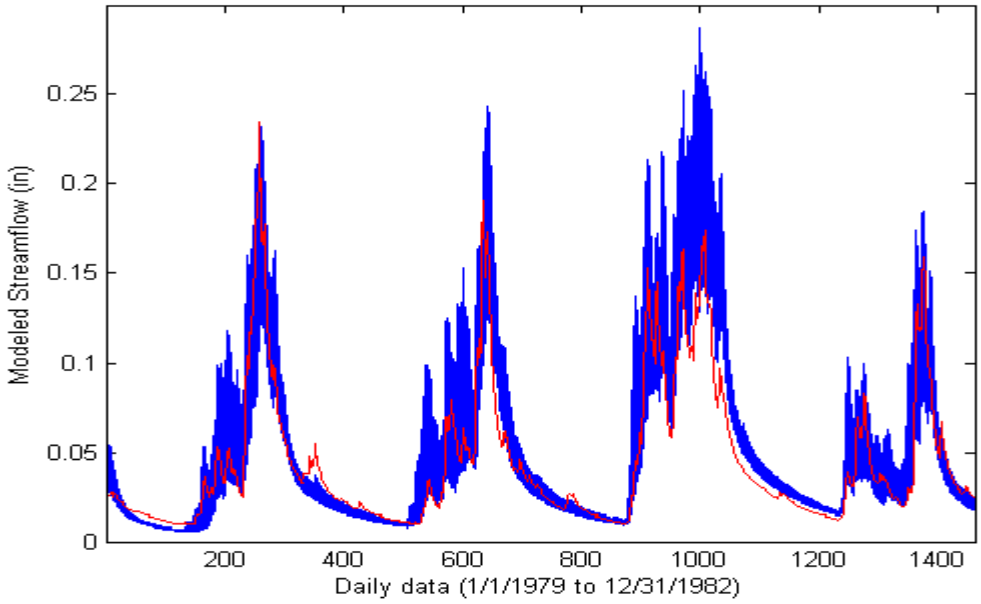


Figure 4.29. Modeled streamflow from raingages (red line) and uncertainty envelope (5% to 95%) for modeled streamflow from modeled rainfall (region in blue).

A possible reason why there are brief periods when the runoff modeled with NARR data are outside the region of uncertainty is that, although the uncertainty intervals employed (5% and 95%) are suitable for theoretical continuous PDFs when the whole domain of the random variable is considered, they are not suitable in the domain of the discrete MOJ-PDF analysis where the knowledge is limited by the available observations. The issue might be partially resolved by increasing the confidence intervals in the observed rainfall (to 0% and 100%). A second approach might be to adjust theoretical PDFs so 5% and 95% confidence intervals are relevant to a broader domain. It is likely that both of these approaches would generate a larger uncertainty region that captures those events currently outside the range. Additional options for improvement are application of the same correction methodology to each of the HRUs separately, rather than to the average rainfall depth for the whole basin, or the use of mobile data bins centered at the value of interest when estimating the individual PDFs instead of bins at regular percentiles. Both of these last two aspects may increase the precision of the results. Finally, the calibration method used, which is based on selecting the percentile that closes the gap between modeled and observed cumulative rainfall, implies non-symmetrical random sampling around the calibrated value during uncertainty estimation. Symmetrical random sampling might be achieved by using the median value multiplied by a correction factor that closes the gap.

5. SUMMARY AND CONCLUSIONS

The objective of this study was to investigate establishing a short- to medium-term hydrometeorological forecasting system for use in Mexico in the spite of the limited data available with which to parameterize the hydrological model required in such a system. The *Río Grijalva* basin was selected as the study basin because of its importance to Mexico's hydropower production network, and because the meteorological phenomena to which it is exposed and the restricted data available for hydrological model parameterization are similar to many other basins in the region. Consequently, a successful study here would bode well for similar future success elsewhere in other Mexican river basins.

Although meteorological forecasting using regional Numerical Weather Prediction (NWP) models is still in an early stage in Mexico, two operational systems do exist and the candidate has past experience in working with one of them. However, currently the use of hydrological models in Mexico, the other essential component of a hydrometeorological forecast system, is practically none existent. For this reason, the present study focused on implementing, calibrating and applying a hydrological model for the *Río Grijalva* basin, and using this model to investigate the hydrologic interpretation of the model calculated meteorological fields available from the North American Regional Reanalysis (NARR). These NARR fields were therefore considered to be representative of those that will be available when the hydrological modeling

component derived in this study is in due course combined with a regional NWP model to create a real time short- to medium-term hydrometeorological forecasting system for use in Mexico.

The hydrologic model MMS-PRMS was selected for use in this study as a physically-based, distributed hydrological model that is highly accessible and customizable, with valuable features such as modular constitution, computing-platform versatility, time-step versatility, and public-domain documentation and code. MMS-PRMS also comes with a parameterization tool (GIS-Weasel) and a parameter optimization tool (LUCA), both of which are valuable for establishing and improving the model and both were used in this study. As mentioned above, in Mexico hydrological modeling is limited by poor documentation of relevant physical properties (topography, soil, vegetation) and sparse meteorological networks. Therefore, a substantial portion of the present work was focused on establishing the hydrologic model itself. This involves parameterization, implementation of meteorological data, and parameter optimization.

A parameterization method for the MMS-PRMS model was proposed and developed which is based on globally available public datasets. It is anticipated that this method will be of general benefit for users outside the US, where the standard parameterization datasets normally used with MMS-PRMS are not available. The model, having been initially parameterized in this way, was first

interfaced to the forcing available from the local meteorological network. A kriging approach was developed and applied to interpolate these data from meteorological stations on to the regular grid required by the model. The interpolated grid data appeared to capture important climatic features adequately and modeled streamflow was able to capture much of the observed variability, thus suggesting that these novel sparse data parameterization and data implementation methodologies work reasonably well.

Nonetheless, the modeled streamflow based on this initial parameterization showed some deficiencies relative to observed baseflow and peak flows. Consequently, optimization of model parameters was explored using the LUCA system provided with the model, which is based on Shuffle Complex Evolution. The large number of parameters used in the model means that an optimization strategy is required in order to avoid creating unrealistic parameter sets, and physical considerations were used to define the optimization procedure used. Arguably the most important feature of this optimization procedure was the order in which model modules are optimized (i.e. “bottom-up”), and the aspect of the model which benefited most from optimization was modeled baseflow. The order in which modules are optimized and the Objective Function selected seem to be responsible for generating this particular improvement. Peak flows, on the other hand, still showed some deficiencies after optimization. However, optimization of peak flows was not explored further because the Objective

Functions (OF) that can be specified in a LUCA process are not appropriate for enhancing deficiencies in this component of the hydrograph. [Note: a recommendation to LUCA developers is that the system should either allow the use of user-defined OFs or include OFs focused on more diverse aspects of the hydrograph. Were a suitable OF implemented in LUCA, an optimization procedure similar to that used in this study could be followed, but with the module order reversed (i.e. “up-bottom”)].

When the NARR data fields were applied to the optimized model, it became apparent that they heavily underestimate rainfall, and that there is systematic one-day lag in the NARR rainfall estimates. The analysis of the Modeled-Observed Joint - Probability Distribution Function (MOJ-PDF) for seasonal rainfall revealed important characteristics in these biases. Specifically, strong rainfall events tended to be overestimated, while weak events tended to be underestimated. The weakest but most frequent events during the rainy season therefore seem to be responsible for the overall underestimation given by NARR data. In addition, one of the most important features of the biases in rainfall is the presence of heteroscedasticity (their standard deviation is proportional to the magnitude of the rainfall event). Because of this, the application of linear, deterministic, bias-correction methodologies is of limited value. Consequently a probabilistic approach based on the MOJ-PDF was adopted.

When a correction-methodology based on the MOJ-PDF was applied to NARR rainfall estimates, the modeled streamflow showed substantial improvement, and comparison with a second alternative correction methodology based on a Maximum Likelihood Estimator (MLE) showed consistent results. The MOJ-PDF analysis can be extended to estimate the uncertainty in modeled rainfall and the consequent uncertainty in modeled streamflow. The uncertainty envelope estimated in this way contains the benchmark data (i.e. the modeled streamflow calculated from raingage data) most of the time, thus suggesting the methodology used for uncertainty estimation is appropriate.

In summary, the primary novel methods, products and conclusions resulting from this study are as follows:

1. A parameterization method for the MMS-PRMS model was developed based on globally available public datasets and therefore appropriate for use outside the US where the standard parameterization datasets normally used with MMS-PRMS are not available.
2. A kriging approach was developed and applied to interpolate data from the local meteorological network on to the regular grid required by the model.
3. Comparison between observed streamflow and the modeled streamflow obtained using the novel sparse data parameterization and data

implementation methodologies summarized as 1 and 2 above suggest that they work reasonably well.

4. A “bottom-up” parameter optimization strategy was defined and applied based on the LUCA tool which resulted in further improvements in modeled baseflow.
5. When NARR data fields were applied to this optimized model, it became apparent that they greatly underestimate rainfall and there is systematic one-day lag in the data relative to observations.
6. A correction-methodology was created based on the Modeled-Observed Joint - Probability Distribution Function (MOJ-PDF) MOJ-PDF which, when applied to NARR rainfall estimates, resulted in a substantial improvement in the modeled streamflow.
7. The MOJ-PDF analysis was extended to estimate the uncertainty in modeled rainfall and the consequent uncertainty in modeled streamflow.
8. Although the present work only documents important biases in NARR-rainfall for the basin of interest, it is strong recommended that in future potential users should consider the estimating biases in NARR data prior to using them.

REFERENCES

BANDAS (Banco Nacional de Datos de Aguas Superficiales) [CD-ROM], Comision Nacional del Agua (CNA) y Instituto Mexicano de Tecnología del Agua (IMTA), 1998.

DesInventar 6. 2007.
Ref Type: Computer Program

Global Land Cover Characteristics Data Base [CD-ROM], 2006a.

Shuttle Radar Topography Mission [CD-ROM], 2006b.

Baldemar-Mendez, A., Aplicación Hidrológica De Los Radares Meteorológicos, Ph.D. UNAM, Ciudad Universitaria, Mexico, 2005.

Buizza, R. and A. Hollingsworth, Storm Prediction Over Europe Using the ECMWF Ensemble Prediction System, *Meteorological Applications, 9*(3), 289-305, 2002.

Buizza, R. and T. N. Palmer, The Singular-Vector Structure of the Atmospheric General circulation, *J. Atmospheric Science, 52*, 1434-1456, 1995.

Chessa, P. A. and F. Lalaurette, Verification of the ECMWF Ensemble Prediction System Forecasts: A Study of Large-scale Patterns, *Weather and Forecasting, 16*(5), 611-619, 2001.

CONAGUA. Estadísticas del Agua en México 2006. 2006. Comision Nacional del Agua.
Ref Type: Report

Daly, C., R. P. Neilson and D. L. Phillips, A statistical-topographic model for mapping climatological precipitation over mountainous terrain, *J. Appl. Meteor., 33*, 140-158, 1994.

De Roo, A. P. J., B. Gouweleeuw, J. Thielen, J. Bartholmes, P. Bongioannini-Cerlini, E. Todini, P. D. Bates, M. Horritt, N. Hunter, K. Beven, F. Pappenberger, E. Heise, G. Rivin, M. Hils, A. Hollingsworth, B. Holst, J. Kwadijk, P. Reggiani, M. Van Dijk, K. Sattler and E. Sprokkereef, Development of a European flood forecasting system, *Intl. J. River Basin Management, 1*(1), 49-59, 2003.

Duan, Q., S. Sorooshian and V. K. Gupta, Optimal use of the SCE-UA global optimization method for calibrating watershed models, *Journal of Hydrology, 158*, 265-284, 1994.

Duan, Q., S. Sorooshian and V. K. Gupta, Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resources Research, 28*(4), 1015-1031, 1992.

Duan, Q., S. Sorooshian and V. K. Gupta, A Shuffled Complex Evolution approach for effective and efficient global minimization, *J. of Optimization Theory and its Applications, 76*(3), 501-521, 2006.

Gupta, H. V., S. Sorooshian and P. O. Yapo, Towards Improved Calibration of Hydrologic Models: Multiple and Non-Commensurable Measures of Information, *Water Resources Research, 34*(4), 751-763, 1998.

Hay, L. E. 2006.
Ref Type: Personal Communication

Hay, L. E., G. H. Leavesley and M. P. Clark, Use of Remotely-Sensed Snow Covered Area in Watershed Model Calibration for the Sprague River, Oregon, 2006a.

Hay, L. E., G. H. Leavesley, M. P. Clark, S. L. Markstrom, R. J. Viger and M. Umemoto, Step-wise, multiple-objective calibration of a hydrologic model for a snowmelt-dominated basin., *Journal of the American Water Resources Association,* 2006b.

Leavesley, G. H., R. W. Lichty, B. M. Troutman and L. G. Saindon, Precipitation-runoff modeling system--User's manual, 1983.

Leavesley, G. H., P. J. Restrepo, S. L. Markstrom, M. Dixon and L. G. Stannard, The modular modeling system - MMS: User's manual, 1996.

Lin, Y., K. Mitchell, E. Rogers, M. E. Baldwin and G. DiMego, Test assimilations of the realtime, multi-sensor hourly precipitation analysis into the NCEP Eta Model, 1999.

Magaña, V., J. A. Amador and S. Medina, The Midsummer Drought over Mexico and Central America, *Journal of Climate, 12*(6), 1577-1588, 1999a.

Magaña, V., J. L. Pérez, J. L. Vázquez, E. Carrisoza and J. Pérez, Los Impactos de El Niño en México. Capítulo 2: El Niño y el Clima, Dirección General de Protección Civil, 1999b.

Mesinger, F., G. DiMego, E. Kalnay, K. Mitchell, P. Shafran, W. Ebisuzaki, D. Jovic, J. Woollen, E. Rogers, E. Berbery, M. Ek, Y. Fan, R. Grumbine, W. Higgins, H. Li, Y. Lin, G. Manikin, D. Parrish and W. Shi, North American Regional Reanalysis: A long-term, consistent, high-resolution climate dataset for the North American domain, as a major improvement upon the earlier global reanalysis datasets in

both resolution and accuracy, *Bulletin of the American Meteorological Society,* 2006.

Pérez, J. B., Pronóstico Numérico Del Tiempo Para El Valle De México, Master Posgrado En Ciencias De La Tierra, U.N.A.M., 2004.

Quintas, I., *Extractor Rápido de Información Climática* [CD-ROM], Instituto Mexicano de Tecnología del Agua (IMTA), 1996.

Reynolds C.A., Jackson T.J. and Rawls W.J., Estimating Available Water Content by Linking the FAO Soil Map of the World with Global Soil Profile Databases and Pedo-transfer Functions, 1999.

Rodriguez, E., C. S. Morris, J. E. Belz, E. C. Chapin, J. M. Martin, W. Daffer and S. Hensley. An assessment of the SRTM topographic products. Technical Report JPL D-31639, -143. 2005. Jet Propulsion Laboratory, Pasadena, California. Ref Type: Report

Simmons, A. J. and A. Hollingsworth, Some Aspects of the Improvement in Skill of NumericalWeather Prediction, *Quarterly J. Royal Meteorological Society, 128*, 647-677, 2002.

Sorooshian, S. and J. A. Dracup, Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: correlated and heteroscedastic error cases, *Water Resources Research, 16*(2), 430-442, 1980.

Uribe, E. M., Analisis de la Variabilidad de la Precipitacion en Tamaulipas, Graduate U.N.A.M., Ciudad Universitaria, Mexico, 2000.

Uribe, E. M., El Inicio de la Temporada de Lluvias en la Costa Sudoeste de Mexico: Relaciones para su Diagnóstico y Pronóstico, M.C. U.N.A.M., Ciudad Universitaria, Mexico, 2002.

Viger, R. J., S. L. Markstrom and G. H. Leavesley, The GIS Weasel - An Interface for the Treatment of Spatial Information Used in Watershed Modeling and Water Resource Management, 1998.

Westrick, K. J. and C. F. Mass, An evaluation of a high resolution hydrometeorological modeling system for prediction of a cool-season flood event in a coastal mountainous watershed, *J. Hydrometeorol.,*(2), 161-180, 2001.

Westrick, K. J., P. Storck and C. F. Mass, Description and evaluation of a hydrometeorological forecast system for mountainous watersheds, *Weather and Forecasting,*(17), 250-262, 2002.